



CONCISE ENCYCLOPEDIA OF PHILOSOPHY OF LANGUAGE AND LINGUISTICS

ALEX BARBER • ROBERT J. STANTON



CONCISE ENCYCLOPEDIA OF PHILOSOPHY OF LANGUAGE AND LINGUISTICS

This page intentionally left blank

CONCISE ENCYCLOPEDIA OF PHILOSOPHY OF LANGUAGE AND LINGUISTICS

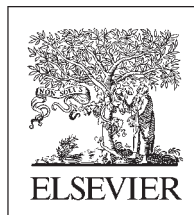
COORDINATING EDITOR

PROFESSOR KEITH BROWN
University of Cambridge
Cambridge, UK

VOLUME CO-EDITORS

PROFESSOR ALEX BARBER
The Open University
Milton Keynes, UK

PROFESSOR ROBERT J. STANTON
University of Western Ontario
London, ON
Canada



Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

Elsevier Ltd., The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

© 2010 Elsevier Ltd.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic, or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publishers.

Permissions may be sought directly from Elsevier's Rights Department in Oxford, UK: phone (+44) 1865 843830; fax (+44) 1865 853333; e-mail permissions@elsevier.com.

*Requests may also be completed online via the homepage
(<http://www.elsevier.com/locate/permissions>).*

Material in this work originally appeared in the *Encyclopedia of Language and Linguistics (Second Edition)*, Ed. K. Brown, Elsevier Ltd., 2006.

Library of Congress Control Number: 2009932541

A catalogue record for this book is available from the British Library

ISBN 978-0-08-096500-0

Printed and bound in the UK

10 11 12 13 14 10 9 8 7 6 5 4 3 2 1

This book is printed on acid-free paper

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

THE EDITORS



Robert J. Stainton was introduced to philosophy of language and to linguistics (specifically in the Systemic Functional tradition) at Toronto's Glendon College. He completed his doctoral work in both areas at MIT. He has been Canada Research Chair in Cognitive Science at Carleton University in Ottawa, and is presently Professor of Philosophy and Associate Dean (Research) at The University of Western Ontario in London, Canada. His books include *Philosophical Perspectives on Language* and *Words and Thoughts*.



Alex Barber is Lecturer in Philosophy at the Open University. He works on the philosophy of language and linguistics as well as on various topics in other areas of philosophy, such as philosophy of mind, epistemology, and more recently, ethics. His published work includes papers on definite descriptions, opacity, speaker versus word reference, and syntactic reality, and various textbooks including *Language and Thought* (Open, 2005). He edited *Epistemology of Language* (Oxford, 2003).

This page intentionally left blank

ALPHABETICAL LIST OF ARTICLES

Subject Classification		xiii
Introduction		xvii
Contributors		xix
<i>A Priori</i> Knowledge: Linguistic Aspects	<i>G Lavers</i>	1
Action Sentences and Adverbs	<i>D Blair</i>	3
Analytic Philosophy	<i>M Beaney</i>	6
Analytic/Synthetic, Necessary/Contingent, and <i>a Priori/a Posteriori</i> : Distinction	<i>R Elugardo</i>	10
Anaphora: Philosophical Aspects	<i>J King</i>	19
Architecture of Grammar	<i>R Kempson</i>	21
Aristotle and Linguistics	<i>P A M Seuren</i>	25
Assertion	<i>M S Green</i>	28
Behaviorism: Varieties	<i>P Poirier</i>	31
Boole and Algebraic Semantics	<i>E L Keenan and A Szabolcsi</i>	36
Causal Theories of Reference and Meaning	<i>A Sullivan</i>	41
Character versus Content	<i>C Spencer</i>	44
Cognitive Science and Philosophy of Language	<i>S Scott</i>	47
Communication: Semiotic Approaches	<i>E Rigotti and S Greco</i>	57
Communication, Understanding, and Interpretation: Philosophical Aspects	<i>D Hunter</i>	64
Comparatives: Semantics	<i>C Kennedy</i>	68
Compositionality: Philosophical Aspects	<i>F J Pelletier</i>	71
Compositionality: Semantic Aspects	<i>G Sandu and P Salo</i>	75
Concepts	<i>E Margolis and S Laurence</i>	78
Conditionals	<i>S Kaufmann</i>	81
Context and Common Ground	<i>H H Clark</i>	85

viii **Alphabetical List of Articles**

Context Principle	<i>R J Stainton</i>	88
Contextualism in Epistemology	<i>H Jackman</i>	94
Conventions in Language	<i>M Kölbel</i>	96
Cooperative Principle	<i>K Lindblom</i>	100
Coreference: Identity and Similarity	<i>Y Huang</i>	107
Counterfactuals	<i>S Barker</i>	109
Creativity in Language	<i>E V Clark</i>	112
Data and Evidence	<i>C Schütze</i>	117
<i>De Dicto</i> versus <i>De Re</i>	<i>E Eaker</i>	124
Default Semantics	<i>K Jaszczolt</i>	128
Definite and Indefinite	<i>B Abbott</i>	131
Definitions: Uses and Varieties of	<i>G Longworth</i>	138
Deflationism	<i>G Marti</i>	141
Deixis and Anaphora: Pragmatic Approaches	<i>K Green</i>	144
Description and Prescription	<i>G Nelson</i>	146
Descriptions, Definite and Indefinite: Philosophical Aspects	<i>G Ostertag</i>	151
Direct Reference	<i>A Sullivan</i>	157
Discourse Representation Theory	<i>J van Eijck</i>	160
Donkey Sentences	<i>P A M Seuren</i>	169
Dthat	<i>D Braun</i>	171
Dynamic Semantics	<i>J Groenendijk and M Stokhof</i>	172
E-Language versus I-Language	<i>G S Rattan</i>	179
Empiricism	<i>J Prinz</i>	183
Empty Names	<i>B Caplan</i>	185
Epistemology and Language	<i>H Jackman</i>	189
Essential Indexical	<i>K Romdenh-Romluc</i>	191
Event-Based Semantics	<i>P Lasersohn</i>	192
Evolution of Semantics	<i>V Evans</i>	196
Evolution of Syntax	<i>B Clark and R Jackendoff</i>	204
Existence	<i>B Caplan</i>	211
Expression Meaning versus Utterance/Speaker Meaning	<i>A Bezuidenhout</i>	214
Expressive Power of Language	<i>D Gil</i>	217
Extensionality and Intensionality	<i>N Oldager</i>	219
Externalism about Content	<i>J Brown</i>	222
Fictional Discourse: Philosophical Aspects	<i>D Davies</i>	227
Figurative Language: Semiotics	<i>F Nuessel</i>	230

Formal Semantics	<i>G Chierchia</i>	243
Formalism/Formalist Linguistics	<i>P ten Hacken</i>	258
Frame Problem	<i>C Viger</i>	265
Functionalist Theories of Language	<i>C S Butler</i>	268
Future Tense and Future Time Reference	<i>Ö Dahl</i>	276
Game-Theoretical Semantics	<i>J Hintikka and G Sandu</i>	279
Generative Grammar	<i>J Rooryck</i>	282
Generative Semantics	<i>J D McCawley and R A Harris</i>	284
Generic Reference	<i>G Carlson</i>	289
Grammatical Meaning	<i>Ö Dahl</i>	293
Holism, Semantic and Epistemic	<i>R Vallée</i>	295
Ideational Theories of Meaning	<i>E J Lowe</i>	299
Identity and Sameness: Philosophical Aspects	<i>H W Noonan</i>	302
Immunity to Error through Misidentification	<i>K Romdenh-Romluc</i>	305
Implicature	<i>J Meibauer</i>	308
Indeterminacy, Semantic	<i>M Hymers</i>	322
Indexicality: Philosophical Aspects	<i>E Corazza</i>	324
Innate Ideas	<i>E Margolis and S Laurence</i>	327
Innate Knowledge	<i>T Simpson</i>	330
Intention and Semantics	<i>S Barker</i>	333
Interpreted Logical Forms	<i>M Montague</i>	336
Irony	<i>S Attardo</i>	341
Language as an Object of Study	<i>V Žegarac</i>	345
Lexical Conceptual Structure	<i>J S Jun</i>	349
Lexical Semantics: Overview	<i>J Pustejovsky</i>	357
Limits of Language	<i>G Priest</i>	365
Linguistic Reality	<i>L Wetzel</i>	368
Linguistics as a Science	<i>B Clark</i>	371
Linguistics: Approaches	<i>N Fabb</i>	378
Linguistics: Discipline of	<i>N Smith</i>	383
Logic and Language: Philosophical Aspects	<i>G Callaghan and G Lavers</i>	398
Logical Consequence	<i>P Blanchette</i>	407
Logical Form in Linguistics	<i>D Blair</i>	410
Lying, Honesty, and Promising	<i>D Owens</i>	413
Mass Nouns, Count Nouns, and Non-count Nouns: Philosophical Aspects	<i>H Laycock</i>	417
Maxims and Flouting	<i>A K Greenall</i>	421

x **Alphabetical List of Articles**

Meaning: Cognitive Dependency of Lexical Meaning	<i>P A M Seuren</i>	424
Meaning: Development	<i>E V Clark</i>	426
Meaning: Overview of Philosophical Theories	<i>R M Martin</i>	433
Meaning: Procedural and Conceptual	<i>D Blakemore</i>	438
Mentalese	<i>F Egan</i>	442
Metalanguage versus Object Language	<i>K Allan</i>	444
Metaphor: Philosophical Theories	<i>M Arseneault</i>	445
Metaphor: Psychological Aspects	<i>R Gibbs</i>	449
Metaphysics, Substitution <i>Salva Veritate</i> and the Slingshot Argument	<i>R J Stainton</i>	456
Modal Logic	<i>M Cresswell</i>	465
Modern Linguistics: 1800 to the Present Day	<i>R Harris</i>	472
Modularity	<i>R Carston</i>	480
Monotonicity and Generalized Quantifiers	<i>R Cann</i>	482
Montague Semantics	<i>T M V Janssen</i>	485
Mood, Clause Types, and Illocutionary Force	<i>K Allan</i>	496
Natural Kind Terms	<i>S Gross</i>	501
Natural versus Nonnatural Meaning	<i>A Barber</i>	505
Naturalism	<i>J Kulvicki</i>	507
Negation: Philosophical Aspects	<i>D DeVidi</i>	510
Negation: Semantic Aspects	<i>R van der Sandt</i>	513
Nominalism	<i>G Klima</i>	520
Nonmonotonic Inference	<i>K Frankish</i>	524
Nonstandard Language Use	<i>A Bezuidenhout</i>	527
Normativity	<i>S Gross</i>	530
Object-Dependent Thoughts	<i>S Crawford</i>	535
Objectivity in Moral Discourse	<i>M Timmons</i>	539
Objects, Properties, and Functions	<i>R J Stainton</i>	545
Ordinary Language Philosophy	<i>P Snowdon</i>	548
Origin of Language Debate	<i>C Neis</i>	551
Paradoxes, Semantic	<i>T Kenyon</i>	557
Philosophy of Linguistics	<i>P Carr</i>	560
Philosophy of Science and Linguistics	<i>C Wearing</i>	566
Plato and His Predecessors	<i>R Bett</i>	569
Plato's <i>Cratylus</i> and Its Legacy	<i>J E Joseph</i>	571
Plurality	<i>P Lasersohn</i>	574
Polysemy and Homonymy	<i>A Koskela and M L Murphy</i>	577

Possible Worlds: Philosophical Theories	<i>D Gregory</i>	580
Pragmatic Determinants of What Is Said	<i>E Borg</i>	582
Predication	<i>S Rothstein</i>	585
Presupposition	<i>P A M Seuren</i>	589
Principles and Parameters Framework of Generative Grammar	<i>R Freidin</i>	596
Private Language Argument	<i>D Byrne</i>	605
Proper Names: Philosophical Aspects	<i>M Reimer</i>	609
Propositional Attitude Ascription: Philosophical Aspects	<i>L Clapp</i>	613
Propositions	<i>S Crawford</i>	617
Quantifiers: Semantics	<i>E L Keenan</i>	623
Radical Interpretation, Translation and Interpretationalism	<i>M Hymers</i>	631
Realism and Antirealism	<i>K Green</i>	634
Reference: Philosophical Theories	<i>A Sullivan</i>	637
Referential versus Attributive	<i>A Barber</i>	644
Relevance Theory	<i>F Yus</i>	648
Representation in Language and Mind	<i>F Egan</i>	656
Rigid Designation	<i>S Predelli</i>	659
Rules and Rule-Following	<i>A Miller</i>	661
Saussure: Theory of the Sign	<i>P Cobley</i>	665
Scope and Binding: Semantic Aspects	<i>A Szabolcsi</i>	677
Semantic Value	<i>J Dever</i>	680
Semantics of Interrogatives	<i>B Sherman</i>	685
Semantics–Pragmatics Boundary	<i>A Bezuidenhout</i>	688
Sense and Reference: Philosophical Aspects	<i>A Sullivan</i>	696
Situation Semantics	<i>V Akman</i>	699
Social Construction and Language	<i>D von Daniels</i>	702
Speech Acts	<i>Y Huang</i>	705
Syncategoremata	<i>G Klima</i>	714
Syntax–Semantics Interface	<i>J-P Koenig</i>	718
Systematicity	<i>P Robbins</i>	728
Tacit Knowledge	<i>G S Rattan</i>	733
Temporal Logic	<i>A Galton</i>	737
Tense and Time: Philosophical Aspects	<i>L N Oaklander</i>	741
Testimony	<i>D Matheson</i>	744
Thought and Language: Philosophical Aspects	<i>E Eaker</i>	747
Transformational Grammar: Evolution	<i>S J Harlow</i>	752

xii **Alphabetical List of Articles**

Truth Conditional Semantics and Meaning	<i>E Lepore</i>	770
Truth: Primary Bearers	<i>S Goldberg</i>	774
Truth: Theories of in Philosophy	<i>M P Lynch</i>	777
20th-Century Linguistics: Overview of Trends	<i>G Graffi</i>	780
Two-Dimensional Semantics	<i>C Spencer</i>	794
Type versus Token	<i>L Wetzel</i>	797
Use Theories of Meaning	<i>G Martí</i>	801
Use versus Mention	<i>C Spencer</i>	803
Vagueness: Philosophical Aspects	<i>R Keefe</i>	807
Verificationism	<i>M Beaney</i>	810
Index		815

SUBJECT CLASSIFICATION

Historical Foundations

Aristotle and Linguistics – P.A.M. Seuren
Boole and Algebraic Semantics – E.L. Keenan & A. Szabolcsi
Modern Linguistics: 1800 to the Present Day – R. Harris
Plato and His Predecessors – R. Bett
Plato's Cratylus and Its Legacy – J.E. Joseph
Saussure: Theory of the Sign – P. Cobley

Key Concepts from Linguistics for Philosophers

Anaphora: Philosophical Aspects – J. King
Creativity in Language – E.V. Clark
Deixis and Anaphora: Pragmatic Approaches – K. Green
Donkey Sentences – P.A.M. Seuren
Dynamic Semantics – J. Groenendijk & M. Stokhof
Event-Based Semantics – P. Lasnik
Evolution of Semantics – V. Evans
Evolution of Syntax – B. Clark & R. Jackendoff
Expressive Power of Language – D. Gil
Future Tense and Future Time Reference – Ö. Dahl
Game-Theoretic Semantics – J. Hintikka & G. Sandu
Generic Reference – G. Carlson
Grammatical Meaning – Ö. Dahl
Lexical Conceptual Structure – J.S. Jun
Lexical Semantics: Overview – J. Pustejovsky
Logical Form in Linguistics – D. Blair
Meaning: Cognitive Dependency of Lexical Meaning – P.A.M. Seuren
Meaning, Procedural and Conceptual – D. Blakemore
Meaning: Development – E.V. Clark
Modularity – R. Carston
Monotonicity and Generalized Quantifiers – R. Cann
Mood, Clause Types, and Illocutionary Force – K. Allen
Nonmonotonic Inference Origin of Language Debate – C. Neis
Plurality – P. Lasnik
Polysemy and Homonymy – A. Koskela & M.L. Murphy

Predication – S. Rothstein
Presupposition – P.A.M. Seuren
Quantifiers: Semantics – E.L. Keenan
Relevance Theory – F. Yus
Scope and Binding: Semantic Aspects – A. Szabolcsi
Semantic Value – J. Dever
Semantics of Interrogatives – B. Sherman
Syntax-Semantics Interface – J.-P. Koenig

Key Concepts in Philosophy of Language

A Priori Knowledge: Linguistic Aspects – G. Lavers
Action Sentences and Adverbs – D. Blair
Analytic / Synthetic, Necessary/Contingent, and *a Priori/a Posteriori* Distinctions – R. Elugardo
Anaphora: Philosophical Aspects – J. King
Assertion – M.S. Green
Causal Theories of Reference and Meaning – A. Sullivan
Character versus Content – C. Spencer
Communication, Understanding, and Interpretation: Philosophical Aspects – D. Hunter
Comparatives, Semantics – C. Kennedy
Compositionality, Philosophical Aspects – F.J. Pelletier
Compositionality: Semantic Aspects – G. Sandu & P. Salo
Concepts – E. Margolis & S. Laurence
Conditionals – S. Kaufmann
Context and Common Ground – H.H. Clark
Context Principle – R. Stainton
Conventions in Language – M. Kölbel
Cooperative Principle – K. Lindblom
Coreference: Identity and Similarity – Y. Huang
Counterfactuals – S. Barker
Creativity in Language – E.V. Clark
De Dicto versus *De Re* – E. Eaker
Default Semantics – K. Jaszczolt
Definite and Indefinite – B. Abbott
Definitions: Uses and Varieties of – G. Longworth
Deflationism – G. Marti
Deixis and Anaphora: Pragmatic Approaches – K. Green

Descriptions, Definite and Indefinite: Philosophical Aspects – G. Ostertag
 Direct Reference – A. Sullivan
 Donkey Sentences – P.A.M. Seuren
 Dthat – D. Braun
 Empty Names – B. Kaplan
 Essential Indexical – K. Romdenh-Romluc
 Existence – B. Caplan
 Expression Meaning versus Utterance/Speaker Meaning – A. Bezuidenhout
 Extensionality and Intensionality – N. Oldager
 Externalism about Content – J. Brown
 Fictional Discourse: Philosophical Aspects – D. Davies
 Figurative Language – F. Nuessel
 Frame Problem – C. Viger
 Generic Reference – G. Carlson
 Grammatical Meaning – Ö. Dahl
 Holism, Semantic and Epistemic – R. Vallée
 Identity and Sameness: Philosophical Aspects – H.W. Noonan
 Immunity to Error through Misidentification – K. Romdenh-Romluc
 Implicature – J. Meibauer
 Indeterminacy, Semantic – M. Hymers
 Indexicality: Philosophical Aspects – E. Corazza
 Innate Ideas – E. Margolis & S. Laurence
 Innate Knowledge – T. Simpson
 Interpreted Logical Form – M. Montague
 Irony – S. Attardo
 Limits of Language – G. Priest
 Linguistic Reality – L. Wetzel
 Logical Consequence – P. Blanchette
 Mass Nouns, Count Nouns, and Non-Count Nouns: Philosophical Aspects – H. Laycock
 Maxims and Flouting – A.K. Greenall
 Mentalese – F. Egan
 Metalanguage versus Object Language – K. Allan
 Metaphor: Psychological Aspects – R. Gibbs
 Modal Logic – M. Cresswell
 Natural Kind Terms – S. Gross
 Natural versus Nonnatural Meaning – A. Barber
 Negation: Philosophical Aspects – D. Devidi
 Negation: Semantic Aspects – R. Van der Sandt
 Nonstandard Language use – A. Bezuidenhout
 Normativity – S. Gross
 Object-Dependent Thoughts – S. Crawford
 Objects, Properties, and Functions – R. Stainton
 Pragmatic Determinants of What Is Said – E. Borg
 Private Language Argument – D. Byrne
 Proper Names: Philosophical Aspects – M. Reimer
 Propositional Attitude Ascription: Philosophical Aspects – L. Clapp
 Propositions – S. Crawford
 Referential versus Attributive – A. Barber
 Representation in Language and Mind – F. Egan

Rigid Designation – S. Predelli
 Rules and Rule-Following – A. Miller
 Semantic Value – J. Dever
 Semantics-Pragmatics Boundary – A. Bezuidenhout
 Sense and Reference: Philosophical Aspects – A. Sullivan
 Speech Acts – Y. Huang
 Syncategoremata – G. Klima
 Systematicity – P. Robbins
 Tacit Knowledge – G.S. Rattan
 Temporal Logic – A. Galton
 Truth: Primary Bearers – S. Goldberg
 Type versus Token – L. Wetzel
 Use versus Mention – C. Spencer
 Vagueness: Philosophical Aspects – R. Keefe

Key Concepts in Philosophy of Linguistics

Architecture of Grammar – R. Kempson
 Compositionality, Philosophical Aspects – F.J. Pelletier
 Compositionality: Semantic Aspects – G. Sandu & P. Salo
 Conventions in Language – M. Kölbel
 Description and Prescription – G. Nelson
 E-Language versus I-Language – G.S. Rattan
 Empiricism – J. Prinz

Linguistic Methodologies and Theories

20th-Century Linguistics: Overview of Trends – G. Graffi
 Behaviorism, Varieties of – P. Poirier
 Cognitive Science and Philosophy of Language – S. Scott
 Communication: Semiotic Approaches – E. Rigotti & S. Greco
 Context Principle – R. Stainton
 Data and Evidence – C. Schütze
 Description and Prescription – G. Nelson
 Discourse Representation Theory – J. Van Eijck
 Dynamic Semantics – J. Groenendijk & M. Stokhof
 Formalism/Formalist Linguistics – P. Ten Hacken
 Functionalism Theories of Language – C.S. Butler
 Game-Theoretic Semantics – J. Hintikka & G. Sandu
 Generative Grammar – J. Rooryck
 Generative Semantics – J.D. McCawley & R.A. Harris
 Language as an Object of Study – V. Žegarac
 Linguistics as a Science – B. Clark
 Linguistics: Approaches – N. Fabb
 Montague Semantics – T.M.V. Janssen
 Principles and Parameters Framework of Generative Grammar – R. Freidin
 Situation Semantics – V. Akman
 Transformational Grammar: Evolution – S.J. Harlow

Philosophical Methodologies and Theories

Analytic Philosophy – M. Beaney

Intention and Semantics – S. Barker
 Ideational Theories of Meaning – E.J. Lowe
 Ordinary Language Philosophy – P. Snowdon
 Nominalism – G. Klima
 Naturalism – J. Kulvicki
 Metaphor: Philosophical Theories – M. Arseneault
 Possible Worlds: Philosophical Theories – D. Gregory
 Reference: Philosophical Theories – A. Sullivan
 Truth Conditional Semantics and Meaning – E. Lepore
 Truth, Theories of in Philosophy – M.P. Lynch
 Two-Dimensional Semantics – C. Spencer
 Use Theories of Meaning – G. Martí
 Verificationism – M. Beaney

Philosophical Applications

A Priori Knowledge: Linguistic Aspects – G. Lavers
 Action Sentences and Adverbs – D. Blair
 Context Principle – R. Stainton
 Contextualism in Epistemology – H. Jackman
 Epistemology and Language – H. Jackman
 Lying, Honesty, and Promising – D. Owens
 Metaphysics, Substitution *Salva Veritate*, and the
 Slingshot Argument – R. Stainton
 Objectivity in Moral Discourse – M. Timmons
 Paradoxes, Semantic – T. Kenyon

Radical Interpretation, Translation, and
 Interpretationalism – M. Hymers
 Realism and Antirealism – K. Green
 Social Construction and Language – D. Von Daniels
 Tense and Time: Philosophical Aspects –
 L.N. Oaklander
 Testimony – D. Matheson
 Thought and Language: Philosophical Aspects –
 E. Eaker

Philosophy of Language and Linguistics: Overviews

20th-Century Linguistics: Overview of Trends –
 G. Graffi
 Cognitive Science and Philosophy of Language –
 S. Scott
 Linguistics as a Science – B. Clark
 Linguistics: Approaches – N. Fabb
 Linguistics: Discipline – N. Smith
 Logic and Language: Philosophical Aspects –
 G. Callaghan & G. Lavers
 Meaning: Overview of Philosophical Theories –
 R.M. Martin
 Philosophy of Linguistics – P. Carr
 Philosophy of Science and Linguistics – C. Wearing
 Transformational Grammar: Evolution – S.J. Harlow

This page intentionally left blank

INTRODUCTION

The entries in this volume have been culled from Elsevier's fourteen-volume *Encyclopedia of Language and Linguistics*, 2nd edition (2006, ed. K. Brown). We have selected the topics that, in our view, are most likely to interest current or aspiring philosophers of language or those concerned with questions at the more philosophical end of linguistics. All but a handful of the entries we have settled on were new to that edition. The majority are from the Philosophy of Language section of the larger work, for which we were sub-editors, but around a quarter are from other sections. We have also been ecumenical in our interpretation of 'are most likely to interest'. Standard items you would expect to find in any list of topics *within* the philosophy of language or the philosophy of linguistics are, we trust, well represented. The volume extends well beyond these bounds, however, because we take the view that in order to do the philosophy of *x*, you need to know something about *x*, and often this means knowing what non-philosophers have to say about *x*. Thus alongside the explicitly philosophical entries are entries on topics that philosophers of language and linguistics are likely at some point to need to know about, despite these topics not being philosophical as such. (Indeed, we have gone so far as to include some entries that we think philosophers of language and linguistics *ought* to know about, whether they realise it or not.) So, for example, we have included entries on the more influential grammars, on comparatives, on what a word is, and so forth. We hope that readers will welcome having entries on such topics ready to hand, and will on occasion find themselves being drawn into reading up on topics they did not necessarily open the encyclopedia to find out about.

A word about the relation between philosophy of language and theoretical linguistics is in order. The last few decades have seen at least two fundamental changes: towards philosophy of language for its own sake, and towards the scientific study of ordinary talk. To understand these changes, it is vital to appreciate two corresponding background features of the philosophical landscape. First, for much of the twentieth century philosophy of language was regarded as the core field of philosophy in the sense that other fields were to be approached through its prism. In a weaker form, this is an old idea. But two schools in the twentieth century took this "linguistic turn" to a new level. This leads to the second background feature, namely the long-standing clash between "scientific" versus "actual usage" camps in the philosophy of language. Rudolph Carnap (1891–1970) and fellow logical positivists (see: 'Verificationism') regarded ordinary language as hopelessly inadequate to the needs of science, and so initiated the development of a new and better suited artificial language. Each sentence in this new language would be tied, definitionally, to the conditions under which it could be confirmed or refuted empirically. Anything not expressible in this improved language would then be dismissed as so much metaphysical nonsense. This revisionist project had been largely discredited by the 1960s, but it has cast a long and often unacknowledged shadow across the whole of philosophy. The opposing camp held that there was nothing wrong with ordinary language. Rather, philosophical problems arise when, to use Ludwig Wittgenstein's phrase, 'language goes on holiday'. Many followers of Wittgenstein (1889–1951), in common with the verificationists he and they opposed, regarded careful reflection on language as a route to overcoming philosophical confusion (see: 'Ordinary Language Philosophy').

The revisionist/scientific versus ordinary-usage dispute is represented in this volume. So too is the language-centric philosophical methodology advocated by both parties to that dispute. However, the ideas and methods of the two camps have gradually lost currency, and the relationship between philosophy of language and

linguistics has undergone a quiet revolution as a result. The focus of the volume lies here. Many theorists seek philosophical understanding of language for its own sake, rather than for the sake of generating a theory of language to serve the needs of other branches of philosophy. Part and parcel of this change has been a greater eagerness among philosophers over the past thirty or so years to have their opinions about language be informed by current approaches in theoretical linguistics. At the very least they are less dismissive than earlier generations were of the potential relevance to philosophy of language of empirical findings. That is, in what has been called the 'New Philosophy of Language', instead of seeing actual usage and scientific methodology as standing in opposition to one another, recent scholars have tended to adopt a scientific attitude to language *as it is*.

Such a philosophical study of language calls for a variety of skills, and is now generally acknowledged to be a collaborative endeavour. This is reflected in the coverage in the present encyclopedia. It is also reflected in the title, the relevant contrast being with *Concise Encyclopedia of Philosophy of Language*, the title of the parallel volume extracted from the first edition of Elsevier's *Encyclopedia of Language and Linguistics* (1993, ed. A. Kasher; the shorter work was edited by P. Lamarque and published in 1997).

Contributors to this encyclopedia have pitched entries at a level that caters to the needs both of relatively uninformed users and of those merely wishing for a sense of where things now stand in the relevant literature. For some topics more than others it seemed reasonable for collateral assumptions to be made about likely readers. For example, anyone looking up 'non-monotonic inference' is probably going to have some knowledge of more basic logical notions or to be willing to look these up. In contrast, many of those looking up the 'use/mention' distinction will have had relatively little exposure to either philosophy or linguistics, and the entry has therefore been written with them in mind.

There is some overlap in the entries, particularly those giving historical reviews of periods or topics. Since we do not envisage many users reading the encyclopedia from beginning to end, we see this duplication as harmless.

Readers looking for the views of a particular person are advised to use the index rather than the table of contents. The latter will make the coverage of key figures seem partial. We have included explicit entries on Plato, Aristotle, and Saussure, for example, but not on, say, Locke, Grice, or Chomsky. The explanation is simple. For the most part, we have taken a topic-based rather than authority-based approach, and the ideas of most key figures have been covered along the way. Locke's theory of language, for example, can be found in the entry on 'Ideational theories of meaning'. Individuals have an entry of their own only if their ideas are not covered in the topic-based entries.

We are hugely grateful to our contributors. All took the job seriously. Some worked to awkward deadlines, while others responded patiently to our occasionally pedantic suggestions. As a group they understood the nature and value of the project and made our job considerably easier than it could have been.

Alex Barber
Rob Stainton

CONTRIBUTORS

Contributors' affiliations are those applicable at the time of submitting their chapters for the *Encyclopedia of Language and Linguistics (Second Edition)*, edited by K. Brown, Elsevier Ltd., 2006.

B Abbott

Michigan State University, East Lansing, MI, USA

V Akman

Bilkent University, Ankara, Turkey

K Allan

Monash University, Victoria, Australia

M Arseneault

University of Wisconsin-Madison, Madison, WI, USA

S Attardo

Youngstown State University, Youngstown, OH, USA

A Barber

The Open University, Milton Keynes, UK

S Barker

University of Nottingham, Nottingham, UK

M Beaney

University of York, York, UK

R Bett

Johns Hopkins University, Baltimore, MD, USA

A Bezuidenhout

University of South Carolina, Columbia, SC, USA

D Blair

University of Western Ontario, London, Ontario, Canada

D Blakemore

University of Salford, Greater Manchester, UK

P Blanchette

University of Notre Dame, Notre Dame, IN, USA

E Borg

University of Reading, Reading, UK

D Braun

University of Rochester, Rochester, NY, USA

J Brown

Bristol University, Bristol, UK

C S Butler

University of Wales, Swansea, UK

D Byrne

University of Birmingham, Birmingham, UK

G Callaghan

Wilfrid Laurier University, Waterloo, Ontario, Canada

R Cann

University of Edinburgh, Edinburgh, UK

B Caplan

University of Manitoba, Winnipeg, Canada

G Carlson

University of Rochester, Rochester, NY, USA

P Carr

Université Paul Valéry, Montpellier, France

R Carston

University College London, London, UK

G Chierchia

Università degli Studi di Milano-Bicocca, Milan, Italy

L Clapp

Illinois Wesleyan University, Bloomington, IL, USA

B Clark

Middlesex University, London, UK

E V Clark

Stanford University, Stanford, CA, USA

H H Clark

Stanford University, Stanford, CA, USA

P Cobley

London Metropolitan University, London, UK

E Corazza

University of Nottingham, Nottingham, UK

S Crawford

Lancaster University, Lancaster, UK

M Cresswell

The University of Auckland, Auckland, New Zealand

Ö Dahl

Stockholm University, Stockholm, Sweden

D von Daniels

University of Leipzig

D Davies

McGill University, Montreal, Canada

J Dever

University of Texas, Austin, TX, USA

D DeVidi

University of Waterloo, Waterloo, Ontario, Canada

E Eaker

University of Western Ontario, London, Ontario, Canada

F Egan

Rutgers University, New Brunswick, NJ, USA

J van Eijck

Centre for Mathematics and Computer Science, Amsterdam, The Netherlands and Research Institute for Language and Speech, Utrecht, The Netherlands

R Elugardo

University of Oklahoma, Norman, USA

V Evans

University of Sussex, Brighton, UK

N Fabb

University of Strathclyde, Glasgow, UK

K Frankish

The Open University, Milton Keynes, UK

R Freidin

Princeton University, Princeton, NJ, USA

A Galton

University of Exeter, Exeter, UK

R Gibbs

University of California, Santa Cruz, CA, USA

D Gil

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

S Goldberg

University of Kentucky, Lexington, KY, USA

G Graffi

University of Verona, Verona, Italy

S Greco

University of Lugano, Lugano, Switzerland

A K Greenall

Norwegian University of Science and Technology, Trondheim, Norway

K Green

Monash University, Melbourne, Australia

K Green

Sheffield Hallam University, Sheffield, UK

M S Green

University of Virginia, Charlottesville, VA, USA

D Gregory

University of Sheffield, Sheffield, UK

J Groenendijk

Universiteit van Amsterdam, Amsterdam, The Netherlands

S Gross

University of Pennsylvania, Philadelphia, PA, USA

P ten Hacken

University of Wales, Swansea, UK

S J Harlow**R A Harris****R Harris**

Oxford, UK

J Hintikka

University of Helsinki, Helsinki, Finland

Y Huang

University of Reading, Reading, UK

D Hunter

State University of New York at Buffalo, NY, USA

M Hymers

Dalhousie University, Halifax, Nova Scotia, Canada

R Jackendoff

Brandeis University, Waltham, MA, USA

H Jackman

York University, Toronto, Ontario, Canada

T M V Janssen

ILLC, University of Amsterdam, Amsterdam, The Netherlands

K Jaszczolt

University of Cambridge, Cambridge, UK

J E Joseph

University of Edinburgh, Edinburgh, UK

J S Jun

Hankuk University of Foreign Studies, Seoul, Korea

S Kaufmann

Northwestern University, Evanston, IL, USA

R Keefe

University of Sheffield, Sheffield, UK

E L Keenan

University of California, Los Angeles, CA, USA

R Kempson

King's College London, London, UK

C Kennedy

Northwestern University, Evanston, IL, USA

T Kenyon

University of Waterloo, Waterloo, Ontario, Canada

J King

University of Southern California, Los Angeles, CA, USA

G Klima

Fordham University, Bronx, NY, USA

J-P Koenig

University at Buffalo, Buffalo, NY, USA

A Koskela

University of Sussex, Brighton, UK

M Kölbel

University of Birmingham, Birmingham, UK

J Kulvicki

Dartmouth College, Hanover, NH, USA

P Lasersohn

University of Illinois at Urbana-Champaign, Urbana, IL, USA

S Laurence

University of Sheffield, Sheffield, UK

G Lavers

University of Western Ontario, London, Ontario, Canada

H Laycock

Queen's University, Kingston, Ontario, Canada

E Lepore

Rutgers University, Piscataway, NJ, USA

K Lindblom

Stony Brook University, Stony Brook, NY, USA

G Longworth

Birkbeck College, University of London, London, England, UK

E J Lowe

University of Durham, Durham, UK

M P Lynch

University of Connecticut, Storrs, CT, USA

E Margolis

Rice University, Houston, TX, USA

G Marti

ICREA & Universitat de Barcelona, Barcelona, Spain

R M Martin

Dalhousie University, Halifax, Nova Scotia, Canada

D Matheson

Carleton University, Ottawa, Ontario, Canada

J D McCawley**J Meibauer**

Universität Mainz, Mainz, Germany

A Miller

Macquarie University, Sydney, NSW, Australia

M Montague

University of California, Irvine, CA, USA

M L Murphy

University of Sussex, Brighton, UK

C Neis

University of Potsdam, Potsdam, Germany

G Nelson

University College London, London, UK

H W Noonan

University of Nottingham, Nottingham, UK

F Nuessel

University of Louisville, Louisville, KY, USA

L N Oaklander

University of Michigan–Flint, Flint, MI, USA

N Oldager

Technical University of Denmark, Lyngby, Denmark

G Ostertag

Nassau Community College, Garden City, NY, USA

D Owens

University of Sheffield, Sheffield, UK

F J Pelletier

Simon Fraser University, Burnaby, British Columbia, Canada

P Poirier

Université du Québec à Montréal, Montréal, Québec, Canada

S Predelli

University of Nottingham, Nottingham, UK

G Priest

University of Melbourne, Melbourne, Victoria, Australia

J Prinz

University of North Carolina, Chapel Hill, NC, USA

J Pustejovsky

Brandeis University, Waltham, MA, USA

G S Rattan

University of Toronto, Ontario, Canada

M Reimer

University of Arizona, Tucson, AZ, USA

E Rigotti

University of Lugano, Lugano, Switzerland

P Robbins

Washington University in St. Louis, St. Louis, MO, USA

K Romdenh-Romluc

University of Nottingham, Nottingham, UK

J Rooryck

Leiden University, Leiden, The Netherlands

S Rothstein

Bar-Ilan University, Ramat Gan, Israel

P Salo

University of Helsinki, Helsinki, Finland

R van der Sandt

Radboud University, Nijmegen, The Netherlands

G Sandu

University of Helsinki, Helsinki, Finland

C Schütze

University of California, Los Angeles, CA, USA

S Scott

Washington University in St. Louis, St. Louis, MO, USA

P A M Seuren

Max Planck Institute for Psycholinguistics, Nijmegen,
The Netherlands

B Sherman

Princeton University, Princeton, NJ, USA

T Simpson

University of Sheffield, Sheffield, UK

N Smith

University College London, London, UK

P Snowdon

University College London, London, UK

C Spencer

Howard University, Washington, DC, USA

R J Stainton

University of Western Ontario, London, Ontario, Canada

M Stokhof

Universiteit van Amsterdam, Amsterdam,
The Netherlands

A Sullivan

Memorial University of Newfoundland, St. John's,
Newfoundland, Canada

A Szabolcsi

New York University, New York, NY, USA

M Timmons

University of Arizona, Tucson, AZ, USA

R Vallée

Université de Moncton, Shippagan, Canada

C Viger

University of Western Ontario, London, Ontario, Canada

C Wearing

University of Western Ontario, London, Ontario, Canada

L Wetzel

Georgetown University, Washington, D.C., USA

F Yus

University of Alicante, Alicante, Spain

V Žegarac

University of Luton, Luton, UK

A

A Priori Knowledge: Linguistic Aspects

G Lavers, University of Western Ontario, London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Immanuel Kant made two divisions between types of knowledge. He distinguished between the *a priori* and the *a posteriori* and also between analytic and synthetic judgments. *A posteriori* claims are the most simple to categorize: They include all empirical knowledge. Everything we learn about the world through our senses falls under this category. *A priori* knowledge is knowledge the ultimate justification of which is independent of experience. However, to call a piece of knowledge *a priori* is not to claim that the knowledge is literally prior to all experience (innate). For instance, mathematical knowledge is widely taken to be *a priori*. That is not to say that no experience is necessary to learn mathematical truths. When one claims that mathematics is *a priori*, this means that experience plays no role in the *justification* of a mathematical proposition. The justification and what led one to believe a proposition (the cause of the belief) need to be carefully distinguished. One may require paper and pencil to convince oneself that a certain mathematical proposition is provable. However, once proven, it is the proof and not the experience of the written proof that is the ultimate justification for the proposition.

A proposition is analytic according to Kant if it is true by virtue of meaning. For instance, the claim that all bachelors are unmarried is an analytic proposition because 'bachelor' just means unmarried man. For Kant and most of the analytic tradition in philosophy, all analytic truths are *a priori*. Because analytic claims are true in virtue of meaning, their justification is nonempirical.

Kant was of course familiar only with Aristotelian logic in which the only logical relation is that of containment between subject and predicate. He therefore equated the property of being true by virtue of meaning with being a logical truth in this sense. In the example above, for instance, it is true because the class of unmarried things includes the class of bachelors.

A proposition is analytic according to Kant therefore if and only if the subject concept is contained in the predicate concept.

Whereas all analytic claims are trivial, a synthetic claim is any claim that genuinely extends knowledge. All empirical claims are of this sort, but Kant argued that there were also nontrivial *a priori* truths. Kant held that the class of synthetic *a priori* truths included both geometry and mathematics. Mathematics is synthetic because, for instance, 12 is nowhere contained in the concepts of 7, addition, and 5. Mathematics is also *a priori* because it does not depend for its justification on experience. To show that (Euclidean) geometry must be *a priori*, Kant provided an argument that he called the 'transcendental esthetic.' Here Kant argued that geometrical relations cannot be learned through experience, because to understand something as located in space, we must have already organized our sensations spatially. That is to say, if our minds did not organize sensations spatially, we could not learn anything concerning the structure of space through sensation. This faculty of ours to organize sensations into a single Euclidean space Kant called our form of spatial intuition. That it is our minds that organize experience into a three-dimensional Euclidean space allows Kant to claim that even though geometrical claims are synthetic, they are nonetheless *a priori*. The structure of space is not learned from experience; it is known through the pure intuition of space. Our pure intuition of space is what allows us to have nontrivial *a priori* knowledge in geometry. Kant also argued that we must have a pure intuition of time because if we did not organize our sensations temporally we could not learn of temporal relations through sensation. It is this pure temporal intuition that Kant believed allows us to have synthetic *a priori* knowledge in mathematics.

Kant thought he had established that the laws of Euclidean geometry were synthetic *a priori* truths about empirical space. Yet, even as Kant was writing the arguments just presented, work was being done on the development of non-Euclidean geometries. Later, Hermann von Helmholtz showed that it is possible to imagine a set of experiences that would

2 A Priori Knowledge: Linguistic Aspects

lead one to believe that space is non-Euclidean. By the end of the 19th century, the geometry of our space was considered an open question. In Bertrand Russell's fellowship thesis, he argued that which geometry applies to our space is an empirical question.

Henri Poincaré took issue with Russell's assertion that the geometry of our space is a straightforwardly empirical question. If, for instance, we construct a large triangle out of light rays and then measure the angles and find that they do not sum to 180 degrees, we cannot yet say that the geometry of space is non-Euclidean. This is because, as Poincaré stressed, we require the further assumption that light travels in a straight line. Poincaré argued that, to preserve the simplicity of Euclidean geometry, we are free to postulate that the path of the light rays is not a straight line. He believed that we are free to hold either that light travels in a straight line and space is non-Euclidean or that space is Euclidean and light does not travel in a straight line. Given this situation, it is incorrect to say that space has a certain geometry. The question of the geometry of space is as meaningful as the question of whether space ought to be measured in inches or centimeters. The various geometries are purely abstract theories that say nothing about empirical space until certain stipulations have been made concerning the types of things that are to count as straight lines.

In 1915, Albert Einstein produced his general theory of relativity. This theory, which asserts that the curvature of space depends on the distribution of matter, is well confirmed. Furthermore, there is no flat space-time theory that makes the same predictions as general relativity. Given this situation, we are no longer free, as Poincaré assumed, to retain Euclidean geometry, come what may. However, Poincaré's point that, without intervening assumptions, geometrical propositions say nothing about empirical space still holds. We need to specify that the straight lines through space-time are the paths of freely falling bodies. It now seems clear that what Kant took to be synthetic *a priori* truths about empirical space are actually false. Hence, Einstein's famous quote: "As far as the laws of mathematics refer to reality, they are not certain; and so far as they are certain, they do not refer to reality."

The reaction in the philosophical community – especially the logical positivists in the early 20th century – to the situation just described was to reject the synthetic *a priori*. All *a priori* truths were taken to be analytic. This was motivated not only by the considerations above but also by the development of modern logic that expanded the class of logical truths significantly. Such a statement as 'if there is someone who knows everyone, then every one is known by at

least one person' can now be shown to be a logical truth, but it is certainly not a case of the subject being included in the predicate. Thus, the category of analytic *a priori* truths is expanded, and the class of synthetic *a priori* truths is eliminated entirely. The logical positivists also relativized the *a priori*. Any statement that we wish to hold as a matter of stipulation gains the status of an *a priori* truth. Definitions of theoretical terms or relations between theoretical terms can be taken to be *a priori*. If the theory that they are part of is modified or abandoned, they are no longer taken to be *a priori*. So, the class of *a priori* propositions is revised as we revise our theories.

According to the positivists, the model of scientific theories is as follows. Scientific theories are composed of a certain class of purely theoretical sentences that are taken to be analytic and *a priori*. Given that these claims are purely theoretical, they make no assertions about things that can actually be observed. This independence from anything observable explains their *a priori* status. These sentences are seen as true in virtue of meaning. However, there will also be a class of sentences that relate this theoretical vocabulary to things that can be observed. These are called 'correspondence rules.' These correspondence rules serve to give an empirical interpretation to the theoretical vocabulary.

This positivistic theory of theories came under heavy attack by W. V. O. Quine in the middle of the 20th century. In particular, Quine attacked the division of sentences into analytic and synthetic. For the positivists, given their rejection of the synthetic *a priori*, the rejection of the analytic/synthetic distinction amounts to a rejection of the *a priori/a posteriori* distinction as well. Quine thought that the difference between analytic and synthetic sentences is a matter of degree, rather than a difference in kind. His view was based on two observations. First, the distinction between observational and theoretical vocabulary is itself a difference of degree. Second, because we can reject highly theoretical sentences on the basis of making certain observations (if the theory as a whole is rejected), then it does not seem reasonable to claim that these sentences are independent of experience. For the positivists, such sentences as 'force equals mass times acceleration' function as a definition of force in Newtonian physics and are thus true in virtue of meaning, and hence, *a priori*. Yet, this claim, according to the positivists, says nothing about the world.

Quine stressed that individual sentences do not have identifiable content, but rather, it is theories as wholes that make assertions about the world. Thus, the definition of force, as part of Newtonian physics, does make an assertion about the world

(in this case, that the world is Newtonian). Quine's view that there are no analytic or *a priori* sentences was widely influential; however, there is now renewed interest in making the distinction between *a priori* and *a posteriori* assertions to better understand how scientific theories function.

Quine's attack on the analytic/synthetic distinction eventually faced empirical scrutiny. Linguists such as Noam Chomsky and Ray Jackendoff have shown that there are empirical reasons to hold that there is an analytic/synthetic distinction. Certain sentences seem to show analytic connections with one another. For instance, upon hearing the sentence 'Jane was convinced to leave' one will assume that Jane decided to leave. An analytic sentence is one whose truth is given by the existence of such a connection. Analytic sentences in this sense are true in virtue of meaning. They are also knowable independent of experience and thus *a priori*. However, that they are knowable independently of experience is not itself knowable *a priori* but known through empirical investigation of natural language. Kant imagined that we could identify *a priori* truths *a priori*. Here we have a case of *a priori* truths that are discovered empirically. On this view *a priori* truths can have no foundational epistemological status.

It had been almost universally believed that all analytic claims were *a priori*. This position is shared by Kant and the positivists (the positivists went further in claiming that all *a priori* claims were analytic). However, on the basis of work by Saul Kripke and Hilary Putnam, a case can be made that certain claims are both analytic and *a posteriori*. That is, there are

claims that are true in virtue of meaning, but are not knowable independently of experience. For instance, part of the meaning of the term 'water' is that it is composed of H₂O. So the sentence 'water is H₂O' is true in virtue of meaning, and thus analytic. However, this sentence is nonetheless a substantial claim about the world and certainly not knowable independently of experience, and thus it is *a posteriori*.

See also: Analytic Philosophy; Analytic/Synthetic, Necessary/Contingent, and *a Priori/a Posteriori*: Distinction; Logic and Language: Philosophical Aspects.

Bibliography

- Helmholtz H (1977). *Epistemological writings*. Lowe M F (trans.). Dordrecht and Boston: D. Reidel.
- Kant I (1929). *Critique of pure reason (1781)*. Smith N K (trans.). New York: St Martin's Press.
- Kant I (1950). *Prolegomena to any future metaphysics (1783)*. Beck L W (trans.). Indianapolis: Bobbs-Merrill.
- Poincaré H (1899). 'Des Fondements de la Geometrie.' *Revue de la Metaphysique et Moral* 7.
- Poincaré H (1900). 'Sur les Principes de la Geometrie.' *Revue de la Metaphysique et Moral* 8.
- Quine W V O (1960). *Word and object*. Cambridge, MA: MIT Press.
- Quine W V O (1961). 'Two dogmas of empiricism.' In *From a logical point of view*. New York: Harper & Row.
- Russell B (1956). *An essay on the foundations of geometry (1900)*. New York: Dover.
- Sarkar S (ed.) (1996). *Science and philosophy in the twentieth century: basic works of logical empiricism* (6 vols). New York: Garland.

Action Sentences and Adverbs

D Blair, University of Western Ontario, London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Action Sentences

Philosophers have a long-standing interest in the distinction between the things that people do and the things that merely befall them. This interest stems from a deeper concern with the concept of intentional action and the nature of human agency. Many distinguished philosophers have sought to illuminate this distinction by looking at how actions are described in natural language and what such descriptions imply about the agents of an action and the role of their intentions in the events described.

It is difficult to get at the distinction in a principled way. The problem is especially acute for sentences like (1):

- (1) Phillip made Sally fall into the well.

Example (1) can be made true either by Phillip's deliberate pushing of Sally into the well or by his having accidentally done so. Likewise, a sentence such as:

- (2) John's hand rose

does not imply that the raising of John's hand was the result of anything that John did. Compare this with:

- (3) John raised his hand

which is true only if John intentionally raised his hand. Example (3), but not (2), implies that the

raising of John's hand was an **action** of his, that he intended to raise his hand and that this intention was the reason why his hand rose when it did. But does this difference in how the truth conditions of (3) might be constituted reflect a corresponding semantic difference? Is (3) actually ambiguous?

I will look at two popular ways of capturing the properties of action sentences, starting with one rooted in philosophical logic, the other more closely related to natural language semantics. One of the main traditions is tied to the logic of the deontic modalities, i.e., the logic of obligation. Just as one might translate sentences of the form α is *obligated to* Φ with a special operator with *It is obligatory for α that α Φ* , so one might capture the agentive sense of a sentence of the form α Φ -ed with *It was intentional of α that α Φ -ed*. Some other proposals for paraphrasing are illustrated in examples (4) and (5).

(4a) John swam the channel.

(4b) John brought it about that he swam the channel.

(5a) Sam brought the pie.

(5b) Sam made it the case that he brought the pie.

One can still wonder whether these paraphrases really capture the agentiveness of these constructions. Looking just at (4a), there are a number of ways in which one can bring about the swimming of the channel, not all of them having to do with one's intentionally swimming the channel. Similar remarks apply to *making it the case that*. Not all of what I can intuitively be said to have made the case originates in something I intended to do.

A notable development in this area, one which appears to escape some of the problems of earlier views, is the work Nuel Belnap, along with several colleagues. They treat

(6) John opened the door

on its agentive reading as equivalent to

(7) John saw to it that the door opened.

Adoption of a special operator *stit* (α sees to it that: p) also allows them to capture some of commonalities between action verbs and imperatives, e.g., the fact that imperatives can only be formed felicitously from action verbs. One might worry, though, about the relations the proposed paraphrases have to the meanings of natural language sentences.

Another, popular way of approaching action sentences comes from Donald Davidson. Rather than rendering action sentences in a notation specifically designed to capture the concept of intentional action, e.g., *stit* operators and the like, one seeks to illuminate the semantic properties of action sentences by looking at what a general theory of meaning for the language containing those sentences tells us.

More specifically, the signature move in Davidson's paper was to posit, alongside the overt arguments of the verb, an additional argument place whose value is an event. One natural way of capturing this thought is to look at ordinary ways of classifying objects. A common noun N predicated of an object x classifies x as an N – a chair, a book, etc. Likewise for verbs, if events are taken to be values of variables just as concrete objects are: a verb V predicated of an event e classifies e as a V-ing, e.g., as a pushing or a swearing, etc. The original formulation of Davidson's proposal, ignoring tense, is this:

(8a) John kissed Mary

(8b) $\exists e[\text{kiss}(\text{John}, \text{Mary}, e)]$

The extra argument place in (8b) is for an event. An early modification of Davidson's view is to make (8b) a bit more complex.

(8c) $\exists e[\text{AGENT}(\text{John}, e) \ \& \ \text{kiss}(e) \ \& \ \text{PATIENT}(\text{Mary}, e)]$

Informally, what this says is that there is a kissing whose agent is John and whose patient is Mary, i.e., there is a kissing of Mary by John.

Davidson's views have been influential, both in inspiring further work and in attracting critical attention. For example, what does one say about sentences like:

(9) I haven't eaten breakfast.

Surely, one does not want an utterance of (9) to mean that there is no past event of my having eaten breakfast. Views taking Davidson's work as a point of departure include the work of Jennifer Hornsby (1980) and George Wilson (1989). Pietroski (1998) and Lombard (1986) discuss some of other criticisms, suggesting modifications.

Adverbs

Adding events to the toolkit of philosophers and philosophically minded linguists also enabled new treatments of adverbs. There is perhaps no single reason why philosophers have been interested in adverbs. But a good deal of what is of interest to philosophers has to do with the metaphysics of events or, in terms that partly overlap with talk of events, occasions, happenings, situations, and states of affairs. For example, a good many philosophers have wanted to acknowledge different classes of events, e.g., states, processes, achievements, and so on. See Vendler (1967), Bennett (1988), and Steward (1997). One way to approach these issues is by looking at how adverbs modify verbs that apply to these different sorts of events.

I will look at two families of theories of adverbs, beginning with an application of Davidson's view

of action sentences. Just as adjectives modify nouns, so adverbs modify verbs. The following analysis is nearly unavoidable for many adjectives:

$$[_{NP} \text{Adj} [N]] \Rightarrow [\text{Adj}(x) \ \& \ N(x)]$$

Thus, the denotation of, e.g., ‘nice dog’ is something that is both nice and a dog. Davidson’s event analysis of action verbs gives a parallel, conjunctive analysis of adverbial modification:

$$[_{VP} \text{Adv} [V]] \Rightarrow [\text{Adv}(e) \ \& \ V(e)]$$

The grammatical combination of an adverb with a verb is to be interpreted as the predication of two predicates about an event. The parallel with nominal modification is preserved. The theory also works for phrasal modification of verbs, e.g., for prepositional phrases such as *in the school*, *at five o’clock*, *after he took the kids to school*, etc. A sentence such as (10a) is represented as (10b):

- (10a) Sam sang loudly in the shower
 (10b) $\exists e[\text{AGENT}(\text{Sam}, e) \ \& \ \text{sang}(e) \ \& \ \text{loud}(e) \ \& \ \text{in the shower}(e)]$

In words: there was a singing by Sam in the shower and it was loud.

Any theory of adverbial modification should show why the following inferences are valid:

- (11) Sally broke the eggs quickly in the sink
 (12) Sally broke the eggs quickly
 (13) Sally broke the eggs in the sink
 (14) Sally broke the eggs
 (15) Sally broke something in the sink.

Example (11) logically implies (12) and (13), both of which imply (14). Additionally, (11) and (13), although neither (12) nor (14), imply (15) by themselves. A large number of adverbial modifiers permit detachment inferences. These inferences are immediate on the event-based theory.

Even with these virtues though, Davidson’s theory does not readily extend to adverbs such as *halfway* or *partly*, e.g.,

- (16) John walked partly to the store
 (17) Alex halfway filled the glass with juice.

It does not seem to make any sense to talk about events that are, for example, halfway in and of themselves, although there are events that are halfway fillings of glasses. It seems as though the adverb forms up a complex predicate with the verb rather than standing alone as a predicate of an event. More seriously though, as easily as the theory handles cases such as *John swam quickly to the shore*, there is no obvious way of extending it to *John allegedly climbed*

the garden trellis. There is no event that is both alleged as well as a climbing. See Larson (2001) and Taylor (1985) for further discussion.

Perhaps the most influential way of treating adverbs is to take them to be quantificational in some respect or as forming up intensional operators (Cresswell, 1986 Lewis, 1975). This helps with sentences such as (18) and (19).

- (18) John probably left in a hurry.
 (19) George frequently left a chocolate on his son’s pillow.

Here, an adverbial appears to modify not the event picked out by the predicate, i.e., the leaving, but the whole proposition, i.e., John’s leaving in a hurry. A theory having only events at its disposal is ill equipped to handle these cases. On the other hand, a theory incorporating propositions and intensions and quantification over possible worlds seems to be in comparably better shape to handle these kinds of contexts.

Further, if one treated adverbs as properties of events, it is much easier to treat sentences that contain negations such as the following:

- (20) Rob cleverly didn’t catch a fish.

Here, one might say that Rob’s not catching a fish has the property of being clever. Another important advantage of a view like this is that it affords a nice treatment of indefinite noun phrases such as *a fish* or *a chocolate* in the scope of an adverb such as *frequent*. This can be seen in sentence (19) or in (21) and (22) below:

- (21) Lisa often found a missing check in the Chairman’s desk
 (22) Alex repeatedly felt an odd tickle on his toe.

Example (21) is perhaps best read as implying that it is often enough the case that Lisa, when looking in the chairman’s desk, found a check that was missing. And (22) is not best interpreted as implying that there is some one odd tickle that is felt repeatedly by Alex. Rather, what the sentence is most naturally understood as saying is that there was a set of odd tickles in Alex’s toe that was repetitive. The quantificational view is further developed by Thomason and Stalnaker (1975) and by McConnell-Ginet (1982).

See also: Logical Form in Linguistics; Speech Acts.

Bibliography

- Anscombe G E M (1957). *Intention*. Oxford: Basil Blackwell.
 Austin J L (1970). ‘A Plea for Excuses.’ In Urmson J O & Warnock G J (eds.) *Philosophical papers*, 2nd edn. Oxford: Oxford University Press.

- Belnap N, Perloff M & Xu M (2001). *Facing the future*. Oxford: Oxford University Press.
- Bennett J (1988). *Events and their names*. Indianapolis: Hackett Publishing.
- Cresswell M J (1986). *Adverbial modification*. Dordrecht: Kluwer.
- Davidson D (1963). 'Action, reasons and causes.' Reprinted in Davidson D (2001) *Essays on actions and events* (2nd edn.). Oxford: Oxford University Press.
- Davidson D (1967). 'The logical form of action sentences.' Reprinted in Davidson D (2001) *Essays on actions and events* (2nd edn.). Oxford: Oxford University Press.
- Ernst T (2001). *The syntax of adjuncts*. Cambridge: Cambridge University Press.
- Higginbotham J (1989). 'Elucidations of meaning.' *Linguistics and Philosophy* 12, 465–517.
- Hornsby J (1980). *Actions*. London: Routledge/Kegan Paul.
- Jackendoff Ray (1972). *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- Larson R (2001). 'The Grammar of intensionality.' In Preyer G & Peters G (eds.) *Logical form and language*. Oxford: Oxford University Press.
- Lewis D (1975). 'Adverbs of quantification.' In Keenan E (ed.) *Formal semantics of natural language*. Cambridge: Cambridge University Press.
- Lombard L B (1986). 'How not to flip the prowler: transitive verbs of action and the identity of actions.' In Lepore E (ed.) *Actions and events*. Oxford: Basil Blackwell.
- McConnell-Ginet S (1982). 'Adverbs and logical form: a linguistically realistic theory.' *Language* 58, 144–184.
- Parsons T (1990). *Events in semantics of English*. Cambridge: MIT Press.
- Pietroski P M (1998). 'Actions, adjuncts and agency.' *Mind* 107, 73–111.
- Ross J R (1970). 'Act.' In Davidson D & Harman G (eds.) *Semantics of natural language*. Dordrecht: Kluwer Reidel.
- Steward H (1997). *The ontology of mind: events, processes, and states*. Oxford: Oxford University Press.
- Taylor B (1985). *Modes of occurrence*. Oxford: Basil Blackwell.
- Thomason R & Stalnaker R (1975). 'A semantic theory of adverbs.' *Linguistic Inquiry* 4, 195–220.
- Vendler Z (1967). *Linguistics in philosophy*. Ithaca: Cornell University Press.
- Wilson G (1989). *The intentionality of human action*. Stanford: Stanford University Press.

Analytic Philosophy

M Beaney, University of York, York, UK

© 2006 Elsevier Ltd. All rights reserved.

'Analytic philosophy' is the name that is given to what is now generally regarded as the dominant philosophical tradition in the English-speaking world and increasingly in the rest of the world. In its widest sense, it might be taken to mean any kind of philosophy in which analysis plays a central role. But there are all sorts of kinds of analysis, and analysis – in one form or another – has been part of philosophy since the time of the ancient Greeks. Sometimes it is characterized in terms of the emphasis placed on clarity, rigor, and careful argumentation, often involving logical formalization. This might distinguish it from some types of 'continental' philosophy (as it is misleadingly called), to which analytic philosophy is frequently seen as opposed. But the best way to understand analytic philosophy is historically, as a movement that had its roots in developments around the turn of the 20th century and evolved in various directions in response to them.

The origins of analytic philosophy lie in the work of Gottlob Frege (1848–1925), Bertrand Russell (1872–1970), G. E. Moore (1873–1958), and Ludwig

Wittgenstein (1889–1951). In many ways, early analytic philosophy can be seen as comprising two main strands, one developing from Frege's logical analyses of number statements and Russell's theory of descriptions, and the other from Moore's concern with conceptual analysis in ethics and epistemology, the two strands synthesised to some extent by Wittgenstein in the *Tractatus* (1921).

Frege's Analysis of Number Statements

Central to Frege's logical analysis of number statements is the idea that number statements are assertions about concepts. A statement such as *Jupiter has four moons* is to be understood, not as predicating of Jupiter the property of having four moons, but as predicating of the concept 'moon of Jupiter' the second-level property 'has four instances,' which can be logically defined (cf. Frege, 1884: § 57). The philosophical significance of this idea can be illustrated by taking the case of negative existential statements (number statements involving the number 0), which have caused problems throughout the history of philosophy. Consider the following statement:

(U) Unicorns do not exist.

This statement is true, so it must clearly have a meaning. But what is it for (U) to have a meaning? If we analyze it in subject-predicate terms, in accord with its surface grammatical form, then ‘unicorns’ is the subject and ‘do not exist’ the predicate. But how can we say anything meaningful if ‘unicorns’ does not refer to anything? If unicorns do not exist, then what is it to which we are predicating the property of nonexistence? Should we posit – as Meinong and the early Russell did – the ‘subsistence’ of non-existent objects to act as the subjects of statements about them? On Frege’s account, however, such a statement is to be understood as asserting something, not about unicorns but about the concept of ‘unicorn.’ (U), in other words, really means what is better expressed as:

(U*) The concept ‘unicorn’ is not instantiated.

Interpreted like this, it can then be readily formalized in modern logic:

(U#) $\neg \exists x Fx$.

On Frege’s view, existence should not be seen as a (first-level) predicate, but instead, existential statements are to be analyzed in terms of the (second-level) predicate *is instantiated*, represented by the existential quantifier. To say that something does not exist is to say that the relevant concept is not instantiated.

Russell’s Theory of Descriptions

A similar strategy is employed by Russell in his theory of descriptions (1905). Consider Russell’s famous example:

(K) The present King of France is bald.

This too might tempt us to suppose that a king of France must somehow exist – or ‘subsist’ – in order to be the subject of such a statement. According to Russell, however, (K) is to be analyzed as (K*), which can then be formalized as (K#):

(K*) There is one and only one King of France, and whatever is King of France is bald.

(K#) $\exists x (Kx \ \& \ \forall y (Ky \rightarrow y = x) \ \& \ Bx)$.

Here there is no commitment implied to any non-existent objects. All that is needed to understand the statement is a grasp of the relevant concepts and logical constants (including the quantifiers). Definite descriptions, such as ‘The present King of France,’ on Russell’s view, are to be ‘analyzed away’: they do not have meaning in isolation (by standing for some object) but only in the context of the sentences in which they appear (sentences that have a complex, quantificational structure).

F. P. Ramsey called Russell’s theory of descriptions a ‘paradigm of philosophy’ (1931: 263), a view shared by both Moore and Wittgenstein. As Wittgenstein put it in the *Tractatus* (1921: 4.0031): “It was Russell who performed the service of showing that the apparent logical form of a proposition need not be its real one.” What both Frege’s and Russell’s logical analyses open up is a gap between grammatical (or apparent logical) form and (real) logical form. On the surface, both (U) and (K) seem to have a simple subject-predicate structure, but their underlying structure is more complex, revealed in (U#) and (K#). Generalizing, this suggests a whole program of analysis, the aim being to uncover the real logical form of the various statements we make, especially those that are philosophically significant or are liable to lead us astray. On Frege’s and Russell’s conception, then, analysis involves the rephrasing of problematic statements into ‘correct’ logical form.

Moore’s Conception of Analysis

Although Moore and Russell are often closely associated, and Moore endorsed Ramsey’s praise of Russell’s theory of descriptions (Moore, 1959: 195), Moore was less concerned than either Frege or Russell with the development of a logical system as a means to ‘correct’ ordinary language. Like the early Russell, in his own early work Moore assumed a crude decompositional conception of analysis. Propositions, for the early Moore, were complex concepts, and their ‘analysis’ simply involved their decomposition into their constituent concepts. But alongside this decompositional conception was a looser conception of analysis as clarification. This comes out in *Principia ethica*, in which Moore talks of ‘analysis’ as involving the distinguishing of questions that are misleadingly conflated, remarking that “the more clearly distinct questions are distinguished, the better is our chance of answering [them] correctly” (1903: 27; cf. vii). His later work is characterized by the careful and detailed – indeed, pedantic – attention he pays to the subtle distinctions of ordinary language in an attempt to resolve certain traditional philosophical puzzles, such as the problem of the external world (Moore, 1959: chapter 7).

Wittgenstein’s *Tractatus*

In the *Tractatus* (1921), Wittgenstein expresses his great debt to Frege and Russell but shares Moore’s view that ordinary language is in perfect logical order as it is and does not need ‘correcting’ (cf. 1921: 5.5563). What was needed was not an ideal

language but an ideal notation, revealing the underlying semantic structure of ordinary propositions no longer obscured by their surface syntactic form. The aim of analysis was to uncover the necessary presuppositions of our use and understanding of ordinary language. Notoriously, on Wittgenstein's early view, the logic of our language required the necessary existence of simple objects, a view he was later to repudiate as part of his general critique of the Tractarian picture theory of language and logical atomism.

Logical and Metaphysical Analysis

The early 1930s marked the heyday of the form of analysis embodied in Wittgenstein's *Tractatus*. During this period an important distinction came to be drawn between what was called 'logical' or 'same-level' analysis and 'metaphysical' or 'reductive' or 'directional' or 'new-level' analysis. The first translates the proposition to be analyzed into better logical form, while the second exhibits its metaphysical presuppositions. Take Russell's theory of descriptions again. In offering (K#) as the analysis of (K), the definite description is 'analyzed away': this is logical analysis, revealing the underlying structure of the statement. But once we have (K#), we must still explain the commitments that remain – to the relevant logical constants and concepts. This may in turn require further analysis to 'reduce' them to things of our supposed immediate acquaintance – such as the 'sense data' to which Russell himself appealed.

Ordinary Language Philosophy

During the 1930s, however, metaphysical analysis came under increasing fire, the assumptions of logical atomism being found unsupportable – in particular, the commitment to such things as simple objects or sense data. But this still left logical analysis – combined now with conceptual or linguistic clarification rather than metaphysical reduction. Wittgenstein himself gave up the idea that Fregean logic provided 'the' logic of ordinary language, but he continued to explore the 'logic' – or what he now tended to call the 'grammar' – of our concepts. In this he influenced a whole generation of philosophers, particularly those dominant in Oxford during the 1950s and 1960s: Gilbert Ryle (1900–1976), J. L. Austin (1911–1960), Paul Grice (1913–1988), R. M. Hare (1919–), and Peter Strawson (1919–). Ryle, for example, explored what he called the 'logical geography' of our mental concepts in *The concept of mind* (1949), in which he introduced the important idea of a 'category-mistake.' According to Ryle, Descartes made a

category-mistake in treating the mind as some kind of spiritual substance, in just the same way as a tourist who asks where Oxford University is in Oxford makes a category-mistake. Austin, even more than Moore, saw philosophical clarification as requiring close attention to the subtle nuances of ordinary language, and his distinction between locutionary, illocutionary, and perlocutionary acts played a key role in the creation of speech-act theory. Grice's conception of conversational implicature, too, might be singled out as influential in the development of speech-act theory and, more generally, in the recognition of the pragmatic alongside syntactic and semantic dimensions of language use.

Strawson is perhaps the best-known representative of the ordinary language movement in 20th-century philosophy. His critique of Russell's theory of descriptions in 'On referring' (1950) is as much a classic as Russell's original 1905 paper. Even though Strawson's *Individuals* (1959) marked a return to metaphysics, it was a form of metaphysics that Strawson called 'descriptive' rather than 'revisionary,' aimed at clarifying our basic conceptual frameworks. In a much later work (1992), Strawson talked of 'connective' rather than 'reductive' analysis as being the methodology of philosophy, illustrating just how far 'analytic' philosophy has evolved since the heyday of logical atomism.

Logical Positivism and the Quinean Tradition

But ordinary language philosophy, rooted in the work of Moore and Wittgenstein, represents only one wing of 20th-century analytic philosophy. As indicated above, the other main wing is the tradition that has its roots in Frege's and Russell's logic and extends through the work of Rudolf Carnap (1891–1970), Alfred Tarski (1901–1983), Kurt Gödel (1906–1978), and the other logical positivists and mathematical logicians, to W. V. O. Quine (1908–2000), Donald Davidson (1917–2003), Saul Kripke (1940–), and others. In this branch of analytic philosophy, there is greater concern with technical matters of logic and with its further development and use in understanding basic semantic concepts such as truth, logical consequence, and necessity. Possible world semantics, originating in Kripke's early seminal papers on modal logic, is perhaps the best example of recent work in this tradition, although Davidson's development of Tarski's work has been crucial in the huge interest that there now is in theories of meaning. On the more informal side, Quine's critique of Carnap's analytic/synthetic distinction has been both influential and

controversial (Quine, 1951). Quine made this critique to show the essential continuity of philosophy and science, with a consequent rejection of the view that there is something distinctive about conceptual analysis. But recent years have seen various defenses – albeit in new contexts – of conceptual analysis, most notably by Frank Jackson (1998).

Analytic Philosophy Today

The two wings of analytic philosophy identified here represent, perhaps, the most important divide within analytic philosophy, a divide that was debated in Richard Rorty's classic collection of papers, *The linguistic turn* (1967). The tension between those who see philosophical progress as lying in ever more sophisticated logical, metaphysical, and scientific theories and those who see the task of philosophy as one of conceptual clarification rooted in ordinary linguistic practices has been a persistent theme throughout the history of analytic philosophy. If there is one insight that might be singled out as lying at the heart of analytic philosophy, then it is this: that our use of language is often 'systematically misleading,' to use Ryle's (1932) phrase. But responses to that insight have varied considerably. Analytic philosophy today is a broad-ranging and complex set of intertwining subtraditions whose unity is constituted less by any distinctive set of doctrines than by their shared history. Understood like this, despite suggestions by some that we have now entered a 'postanalytic' age, analytic philosophy can be regarded as stronger and healthier today than it has ever been in the past.

See also: Descriptions, Definite and Indefinite: Philosophical Aspects; Logic and Language: Philosophical Aspects; Logical Form in Linguistics; Meaning: Overview of Philosophical Theories; Ordinary Language Philosophy; Possible Worlds: Philosophical Theories; Reference: Philosophical Theories.

Bibliography

Austin J L (1961). *Philosophical papers*. Urmson J O & Warnock G J (eds.). Oxford: Oxford University Press.
 Baldwin T (2001). *Contemporary philosophy*. Oxford: Oxford University Press.
 Baldwin T (ed.) (2003). *The Cambridge history of philosophy 1870–1945*. Cambridge: Cambridge University Press.
 Beaney M (1996). *Frege: making sense*. London: Duckworth.

Beaney M (2003). 'Analysis.' In the Stanford encyclopedia of philosophy, online at <http://plato.stanford.edu>.
 Biletzki A & Matar A (eds.) (1998). *The story of analytic philosophy*. London: Routledge.
 Coffa J A (1991). *The semantic tradition from Kant to Carnap*. Cambridge: Cambridge University Press.
 Floyd J & Shieh S (eds.) (2001). *Future pasts: the analytic tradition in twentieth-century philosophy*. New York: Oxford University Press.
 Frege G (1884). *Die Grundlagen der Arithmetik*. Breslau: W. Koebner. [Selections tr. in Beaney M (ed.), *The Frege reader*. Oxford: Blackwell. 84–129.]
 Hacker P M S (1996). *Wittgenstein's place in twentieth-century analytic philosophy*. Oxford: Blackwell.
 Jackson F (1998). *From metaphysics to ethics: a defence of conceptual analysis*. Oxford: Oxford University Press.
 Moore G E (1903). *Principia ethica*. Cambridge: Cambridge University Press.
 Moore G E (1959). *Philosophical papers*. London: George Allen and Unwin.
 Passmore J (1966). *A hundred years of philosophy* (2nd edn.). London: Penguin.
 Quine W V O (1951). 'Two dogmas of empiricism.' *Philosophical Review* 60, 20–43.
 Ramsey F P (1931). *The foundations of mathematics*. London: Routledge.
 Reck E H (2002). *From Frege to Wittgenstein*. New York: Oxford University Press.
 Rorty R (ed.) (1967). *The linguistic turn*. Chicago: University of Chicago Press.
 Russell B (1905). 'On denoting.' *Mind* 14, 479–493.
 Russell B (1919). *Introduction to mathematical philosophy*. London: George Allen and Unwin.
 Ryle G (1932). 'Systematically misleading expressions.' *Proceedings of the Aristotelian Society* 32, 139–170. [Reprinted in Rorty (1967), 85–100.]
 Ryle G (1949). *The concept of mind*. London: Penguin.
 Soames S (2003). *Philosophical analysis in the twentieth century* (2 vols). Princeton: Princeton University Press.
 Strawson P F (1950). 'On referring.' *Mind* 59, 320–344.
 Strawson P F (1959). *Individuals*. London: Methuen.
 Strawson P F (1992). *Analysis and metaphysics: an introduction to philosophy*. Oxford: Oxford University Press.
 Stroll A (2000). *Twentieth-century analytic philosophy*. New York: Columbia University Press.
 Urmson J O (1956). *Philosophical analysis: its development between the two world wars*. Oxford: Oxford University Press.
 Wittgenstein L (1921). *Tractatus logico-philosophicus*. Pears D F & McGuinness B (trans.). London: Routledge. [Originally published in German, 1921; originally translated into English by Ogden, C. K., 1922, London: Routledge.]
 Wittgenstein L (1956). *Philosophical investigations*. Anscombe G E M (trans.). Oxford: Blackwell. [3rd edn., 1978].

Analytic/Synthetic, Necessary/Contingent, and a *Priori/a Posteriori*: Distinction

R Elugardo, University of Oklahoma,
Norman, USA

© 2006 Elsevier Ltd. All rights reserved.

Three distinctions – necessary/contingent, *a priori/a posteriori*, and analytic/synthetic – mark three different contrasts. The *a priori/a posteriori* distinction marks an epistemological contrast between two different ways of knowing some proposition or two different ways of being justified in believing a proposition. The necessary/contingent distinction marks a modal contrast between different propositions. The analytic/synthetic distinction marks a semantic contrast between different sentences. Until the work of Immanuel Kant, the prevailing view was that either every truth is strictly necessary, *a priori*, and analytic, or it is contingent, *a posteriori*, and synthetic. Suffice it to say that things have gotten a lot more complicated since Kant. Many philosophers are even dubious about the reality of some or all of these distinctions. In this article, I will sketch the main ideas behind each dichotomy and note some points of controversy.

Necessary/Contingent Distinction

Necessity and contingency are said to be metaphysical properties of propositions, primarily *de dicto* modality. (For some general sources on the metaphysics of modality, see: Driver, 2002; Gendler and Hawthorne, 2002; Linsky, 1977; Loux, 1979; Lycan, 1994; Melia, 2003.) A proposition *P* is necessarily true just in case *P* is true and would have been true no matter how things might have been – the world could not have been such as to render *P* false. *P* is necessarily false just in case *P* is false, and it would have been false no matter how things might have been – the world could not have been such as to render *P* true. (Notice that we defined *de dicto* necessity in terms of possibility and negation: necessarily *P* =_{defn} ¬Possibly ¬*P*. We could just as well have defined *de dicto* possibility in terms of necessity and negation: possibly *P* =_{defn} ¬Necessarily ¬*P*. There is a sense, then, in which one cannot give a definition of an alethic modality that goes beyond this circle.) A proposition *P* is contingently true/false just in case *P* is true/false but could have been false/true (if the world had been otherwise in certain respects). The necessary/contingent distinction is therefore jointly exhaustive and mutually exclusive of every proposition that has a truth-value.

Sentences (1a), (2a), (3a), and (4a) are examples of necessary truths:

- (1a) $12 \times 2 = 24$.
- (2a) Gold has atomic number 79.
- (3a) Biological parents have at least one offspring.
- (4a) All dogs are dogs.

Sentences (1b), (2b), (3b), and (4b) are necessary falsehoods:

- (1b) $12 \times 2 = 25$.
- (2b) Gold is not an element.
- (3b) Some biological parents have no offspring.
- (4b) Some dogs have fleas and no dog has any fleas.

Sentences (1c) and (2c) are contingent propositions:

- (1c) Al Gore won the 2000 U.S. Presidential Election.
- (2c) Water freezes below 32°F.

Although (1c) is false, it might have been true if only Gore had received more electoral votes than Bush from the state of Florida. By the same token, although (2c) is true, it might have been false. After all, the laws of physics and chemistry might not have held to the extent and degree that they do. If they hadn't, then it would not have been true that water freezes below 32°F.

Not all necessary truths are of the same kind. For example, (2a) is a metaphysically necessary truth – it is metaphysically impossible that the atomic number of gold not be 79, i.e., nothing could possibly be gold unless it is an element whose atomic number is 79. Mathematical Platonists who regard numbers as mind-independent abstract objects view (1a) as being a metaphysically necessary truth. According to them, the number 24 could not possibly have been distinct from the number that is the product of 12 and 2, which is the number 24 itself, since nothing can possibly be diverse from itself. The basic idea is the same in both cases: certain statements about a mind-independent reality must be true because of certain metaphysical facts about that reality (as opposed to facts about how we think about it).

Sentence (3a) is said to be a conceptually necessary truth because the concept of being a biological parent includes, as one of its conceptual components, the concept of having a biological offspring. Sentence (3a) is not so much a truth about the world as it is a truth about the concepts we use in thinking about the

world. On this view, conceptual necessity is determined solely by the inferential relations that are (allegedly) constitutive of the contents of concepts that are so related. If lexical concepts have internal definitional structure, then (3a) is true by definition and thus, for that reason alone (according to philosophers like A. J. Ayer), it is a necessary truth (see Ayer, 1946).

Sentence (4a) is an example of a logically necessary truth, since it is an instance of the tautological schema 'All Fs are Fs,' which admits of no possible counterexample. A counterexample to that schema would be a case involving some F that was simultaneously non-F, which is a flat-out contradiction and thus impossible. Because they are not made true by any metaphysical facts about the world, logically necessary truths aren't metaphysically necessary. Nor are they conceptually necessary, since they are not made true by any facts about the (nonlogical) concepts we use to think about the world. They are necessarily true simply by virtue of their logical form.

The *a Priori*/a *Posteriori* Distinction

The *a priori*/a *posteriori* distinction is an epistemological distinction: it characterizes the way a proposition is known or the way a person is justified in believing the proposition. (For more discussion on the *a priori*/a *posteriori* distinction, see: Boghossian and Peacocke, 2000; Bonjour, 1998; Casullo, 1999; Casullo, 2003.) As Kant presents it, *a priori* knowledge is knowledge that one has independently of all possible sense experience. What makes *a priori* knowledge *a priori* for Kant is its universality – it is knowledge of that which constitutes the formal constraints on *all* possible human experiences – and its necessity. For Kant, knowing that every event has a cause is a case of *a priori* knowledge. By contrast, *a posteriori* knowledge is only possible through experience or introspection. For example, knowing that several hurricanes caused massive destruction throughout Florida in 2004 is a case of *a posteriori* knowledge; you couldn't know that proposition is true merely on the basis of rational reflection.

The evidential source of *a priori* knowledge is neither sensory perception, nor introspection, nor memory. For you to be justified *a priori* in holding a belief, it is enough that you rightly recognize the truth of your belief given only your conceptual grasp or understanding of the content of your belief. Your reason for belief has to emerge from rational reflection or from rational insight, assuming that you have a nonempirical faculty of intellection. *A priori* knowledge may require experience to provide the relevant concepts, but, given possession of the concepts, no further role for experience is required for having

a priori knowledge, except insofar as it can play a preservative role in the transmission of *a priori* justification from one source to another, e.g., testimony.

If *a priori* knowledge is genuine, then propositions (5a), (6a), and (7a) are examples of *a priori* knowable truths:

- (5a) The relation *x is taller than y* is asymmetrical and transitive.
- (6a) Whatever is colored is extended.
- (7a) Whatever is red all over is not blue.

In each case, you can know the truth of the proposition simply by grasping its content, reflecting on what would have to be the case for it to be true, and then coming to recognize that the proposition must be true given your understanding of its content. Empirical observation plays no justificatory role in one's being warranted in believing (5a), (6a), and (7a). Knowledge of propositions (1a), (3a), and (4a) above would also be *a priori*.

By contrast, knowledge of (5b), (6b), and (7b) is *a posteriori*, since one's intellectual grasp of their contents is insufficient to secure epistemic warrant; moreover, one needs empirical evidence in order to be justified in believing them:

- (5b) Yao Ming (the famous Chinese NBA basketball player) is taller than Danny DeVito (the popular film actor and comedian).
- (6b) That car has a black exterior and is compact.
- (7c) Some of these cardboard shapes are painted red on the facing side.

Whether there really is *a priori* knowledge is a matter of great debate. Rationalists and Kantians contend that some knowledge is *a priori*, while Quineans, who adopt a naturalistic approach to epistemology, deny the reality of *a priori* knowledge. (For more on this topic, see: Boghossian and Peacocke, 2000; Bonjour, 1998; Casullo, 1999, 2003; Devitt, 1994; Moser, 1987.)

It is quite natural to think that all and only *a priori* knowable propositions are necessary truths. For example, A. J. Ayer presents the following argument for their identification in Ayer (1946). First, if a proposition is *a priori*, then it can be known to be true independently of any experience of the world. But if that's true, then the truth of an *a priori* proposition must not depend on any contingent features of the world; otherwise, one would have to determine whether those features actually obtained in order to know whether the proposition is true – and then it would be an *a posteriori* proposition. Thus, an *a priori* proposition must be true no matter how

things might have turned out in the world, which is the mark of a necessary truth. Second, if a proposition is a necessary truth, then it is true independently of the way things are. In which case, sense experience cannot play any role in explaining how we can know that proposition since experience can only furnish information about how things actually are, and not about how things must be. So, a proposition is *a priori* if it is a necessary truth.

Although Ayer's arguments contain some questionable steps, his arguments are not wildly implausible. More importantly, they set up a major challenge to his empiricism. If, as he and Carnap (1956) claim, mathematical and logical truths are necessary, and hence *a priori*, then we can have nonempirical knowledge; but then we can have some nonexperiential access to certain substantive facts about the world. That conflicts with Ayer's and Carnap's view that sentences, including sentences that express the contents of our knowledge, can have meaning only by being verifiable in principle. Ayer and Carnap thought they could solve this problem by explaining an *a priori* and necessary truth in terms of a linguistic notion of analyticity, namely, a sentence that is true by definition (Ayer) or by linguistic convention (Carnap). The idea is that any sentence that is true by definition must be necessary because, contrary to appearance, it contains no information about the world anyway. At the same time, it must also be *a priori*, because to know the proposition that such a sentence expresses is simply to know the definitional meanings of the terms that make up the sentence, which is enough to know that the sentence is true; experience plays no additional justificatory role. As we shall see later, Quine argued that the Ayer-Carnap gambit ultimately fails. Let us turn, though, to Saul Kripke's view on the connection between the necessary/contingent and the *a priori/a posteriori*.

Kripke on the Necessary *a Posteriori* and the Contingent *a Priori*

As noted in the last section, the prevailing view among philosophers until the 1970s was that all and only necessary truths are *a priori* and that all and only contingent truths are *a posteriori*. Most philosophers have now abandoned the view, thanks largely to Kripke (1972/1980). According to Kripke, metaphysical necessity is not a criterion for the *a priori* since they track entirely different kinds of properties (alethic modal properties and epistemic properties, respectively) of different kinds of things (propositions/sentences and ways of knowing a proposition, respectively). So it should come as no surprise that both can sometimes come apart.

Kripke's example of a sentence that expresses a necessary *a posteriori* truth is 'Water is H₂O.' According to him, the terms 'water' and 'H₂O' are each **rigid designators**: they refer to their bearers in every possible world. Thus, 'water' rigidly designates water, and 'H₂O' designates rigidly the chemical substance that is composed of H₂O molecules. Given that water is chemically composed of H₂O molecules, the two terms co-refer in every possible world. There is no possible world in which something is identical to H₂O but isn't water, so 'Water is H₂O' is true in every metaphysically possible world. It therefore expresses a metaphysically necessary truth.

According to Kripke, that water is H₂O is a *posteriori* knowable. First, it is *a priori* and necessarily true that a substance's chemical composition is essential to it. Hence, anything that has the same chemical nature as water must of necessity be identical to water. However, we can't deduce anything about the actual particular nature of water from the concept of its being a chemical substance or from the concept of its having an empirical nature that is essential to water. We have to do some empirical investigation instead. Only when our best scientific theories have confirmed that the chemical nature of water is H₂O can we then be justified in believing that water is H₂O. Therefore, our knowledge that water is H₂O, which is a necessary truth, is *a posteriori*. In general, all true scientific identities are, for Kripke, necessary *a posteriori* truths. If Kripke was correct, then a proposition's being a metaphysically necessary truth doesn't entail that it is an *a priori* truth. (Some of Kripke's followers deny that necessarily true scientific identities are *a posteriori* knowable; see Salmon, 1993.)

Kripke also holds that some contingent propositions are *a priori*. He gives the example of a speaker who introduces the measurement term 'one meter' to pick out the actual length of a certain stick kept in the science museum in Paris at time *t* (call the stick 'Stick S'). The speaker stipulates, "Let 'one meter' designate the length of Stick S." In Kripke's example, the speaker uses 'the length of Stick S at *t*' to designate the actual length that S has at time *t*, (where *t* is the time of the speaker's utterance of her stipulation). The description used determines the reference of 'one meter' but not its meaning. Although 'one meter' rigidly designates a certain length in all possible worlds, which in the actual world happens to be the length of Stick S at *t*, 'the length of S at *t*' does not designate anything rigidly. For there are possible worlds in which, when heat is applied to Stick S at *t*, S is not one meter long at *t*. Hence, the stipulative definition, properly interpreted, does not say that 'one meter' is to be synonymous with 'the length of

S at t.' Rather, says Kripke, the speaker fixes the reference of 'one meter' by stipulating that 'one meter' is to be a rigid designator of the length that is the actual length of Stick S at t. Imagine now that the speaker goes on to assert, 'Stick S is one meter long.' According to Kripke, the sentence 'Stick S is one meter long' is true but not necessarily true, since S could have had a different length at t under appropriate stresses and strains, heatings or coolings. Consequently, the proposition expressed is contingent rather than necessary.

Kripke next argued that, since the actual length of Stick S was used to fix the reference of 'one meter,' the speaker knows automatically and without empirical evidence that S is one meter long. The speaker's very act of fixing the term's reference constitutes her non-empirical justification for her knowledge of said proposition; she does not have to appeal to any experience in order to know that the proposition she expressed is true. This is something, then, that the speaker who introduced 'one meter' in that way knows *a priori*. Thus, on Kripke's view, the speaker's utterance of 'Stick S is one meter long' expresses a contingent *a priori* truth. If Kripke is right, then a proposition's being a contingent truth doesn't entail that it is an *a posteriori* truth. (Not everyone agrees that Kripke's example is an example of a contingent *a priori* truth; see: Casullo, 1999; Moser, 1987; Soames, 2003.)

The Analytic/Synthetic Distinction

Immanuel Kant, who first coined the analytic/synthetic distinction in Kant (1781/1998), defined an analytic judgment as one whose predicate-concept is contained in its subject-concept. Analytic judgments are 'explicative' in the sense that they add nothing new through the predicate to the subject, e.g., 'all bodies are extended.' A synthetic judgment is one whose predicate-concept is not contained in its subject-concept. Synthetic judgments are 'ampliative' in the sense that they add information to the subject concept that is new and that no analysis could possibly extract from it, e.g., 'all bodies are heavy.'

Although Kant held that all analytic truths are *a priori* and necessary, he didn't think that all *a priori* truths are analytic in his sense, e.g., truths of arithmetic. For instance, according to Kant, in judging that 12 is the sum of 7 and 5, one does not think of 12 when one thinks of the sum of 7 and 5. To affirm that 12 is the number that 7 and 5 sum to, one must also imagine certain changes in time, namely, the successive addition of units, which (for Kant) was an intuition of time. Consequently, the judgment that 12 is the sum of 7 and 5 is ampliative and

is thus synthetic. It is also *a priori*, since the judged proposition is known independently of sense experience.

As many commentators have pointed out, Kant's characterization of the analytic/synthetic distinction excluded judgments that appear to be analytic even though they are not of the subject-predicate form, e.g., 'If Mary *persuaded* Susan to attend college, then Susan *intends* to attend to college.' A more inclusive account was therefore needed. (For a discussion of the history of the analytic/synthetic distinction, see: Coffa, 1991; Pap, 1958.)

Metaphysical Analyticity

Since Kant, philosophers have viewed the analytic/synthetic distinction as a semantic distinction between sentences rather than as a distinction between judgments of the subject-predicate form. The standard definition of analyticity is this: *A sentence is analytic if and only if it is true solely by virtue of the meanings (in other words, definitions) of its constituent terms.* Any true sentence that is not analytic is *synthetic*. For example, sentence (8) below is analytic because its truth-value is supposedly fixed entirely by its meaning and structure. By contrast, sentence (9) is synthetic because its truth-value is not fixed by its meaning and form alone; it also depends on how things are in the world it describes.

(8) All vixens are female foxes.

(9) All vixens are nocturnal predators.

Following Boghossian (1997), I will call the standard definition of analyticity 'metaphysical analyticity', since it is a metaphysical thesis about what constitutes a sentence's having the truth-value that it has. (There are some excellent out-of-print anthologies on the analytic/synthetic distinction: Sleigh, 1972; Sumner and Woods, 1969. Pap, 1958, is still one of the best sources on this topic even though it is dated.)

By our definition, an analytic sentence owes its truth-value entirely to its meaning and form irrespective of any worldly facts. Consequently, no such sentence can undergo a change in truth-value, from truth to falsehood, unless it also undergoes a prior meaning-change. If the sentence were to undergo a shift in meaning, then it would not be expressing the same true proposition as before, since any meaning-change would necessitate a change in the proposition it expresses. Therefore, no analytic sentence could ever be empirically refuted (assuming that there is no change in its meaning).

In his famous 1951 study, W. V. O. Quine argued that no sentence is immune to error (see also

Quine, 1991). On the contrary, every sentence is open to possible refutation without any prior change of meaning. Thus, Quine denied that there are any analytic sentences in the metaphysical sense defined above. Let us, then, turn to his argument.

Quine's Belief-Revisability Argument

To set the stage for Quine's argument, I will briefly review the main points at the end of Quine (1951). In that article, Quine took aim against the traditional empiricist's 'dogma' that each meaningful sentence has its own experiential confirmation conditions apart from any background theory. For the empiricists, most notably Ayer (1946) and Carnap (1956), all substantial knowledge comes from experience. For them, confirming a sentence is just a matter of having the appropriate sensory experiences, ones that would confirm the sentence to some degree of probability. Having the appropriate sensations is enough, then, to verify a single sentence; you don't need to assume the truth of any other sentence in order to verify it.

Quine rejected that last claim, since he agreed with Pierre Duhem that our beliefs make contact with experience as a whole, not individually; see Duhem (1914/1954). Call that the Quine-Duhem Thesis. Verifying a single sentence presupposes, then, a massive background of beliefs and default assumptions, which brings us to Quine's Confirmation Holism: *the evidential status of any one belief is fixed by its evidential relations to other beliefs, and their evidential status is in turn fixed by their evidential relations to many other beliefs, and so on for all of one's beliefs*. Thus, no single experience or group of experiences could ever refute or confirm a sentence independently of its evidential connections to many other sentences.

It follows from the Quine-Duhem Thesis and Quine's Confirmation Holism that revising any one belief will necessitate a wholesale revision of one's belief-system. Quine presented his famous "web of belief" metaphor in Quine (1951) to explain this idea more fully. The beliefs that form the center of the 'web' are the ones that are almost impossible to give up (classical truths of logic, mathematical truths). The ones at the extreme ends of the web are the ones that are very easy to give up (observational beliefs), and the ones in between are relatively easy/hard to give up (scientific theoretical beliefs). Still, any belief can be rationally given up or revised for the right price and, by the same token, any belief can be held tenaciously, come what may, but at a cost. What to revise will depend on how much it will cost to preserve the system with respect to its coherence, simplicity, predictive power, and elegance.

Belief-change, then, is solely a function of practical utility: change those beliefs that will require the least amount of changes in the rest of your belief-system while maximizing coherence, simplicity, etc.

Quine's Belief-Revisability Argument against the reality of analytic truths consists of just three premises. First Premise: Quine's Confirmation Holism – there are sentences that you hold true and sentences that you hold false, but in each case the support of your belief is a complex matter of the holistic evidential relations your sentence bears to many other sentences. Second Premise: if Quine's Confirmation Holism is true, then no belief or sentence is completely immune to revision; every sentence can be rejected under pressure from empirical evidence plus a concern for overall coherence. (Putnam, 1983, is an interesting objection to this premise.) Third Premise: if there are any analytic sentences as traditionally conceived, then they must be unresponsive to the world's input, and so immune to revision. Therefore: there are no analytic sentences as traditionally conceived.

Frege-Analyticity

A different conception of analyticity can be found in Frege (1968) and is one that is also discussed in Quine (1951): *A sentence is analytic if and only if either (i) it is a logical truth, or (ii) it can be converted into a logical truth by substitution of synonymous expressions, salva veritate, and formally valid inferences*. Clause (i) mentions a logical truth, which, according to Quine, is any true sentence that remains true under every uniform semantic interpretation and reinterpretation of its component, nonlogical, descriptive terms. Consider, for example, sentence (10):

(10) Every brother is a brother.

The nominal 'brother' is the only nonlogical descriptive term occurring in (10); (10) is true and it remains true under every possible, uniform, (referential) interpretation of the two occurrences of 'brother.' It is therefore a logical truth and thus analytic by clause (i).

Consider now sentence (11):

(11) Every brother is a male sibling.

Although (11) is true, it is not a logical truth, since one could consistently reinterpret 'brother' and 'male sibling' differently by assigning different extensions to them, resulting in a reading under which (11) is false. Still, it can be transformed into a logical truth by replacing 'male sibling' in (11) with its synonym 'brother,' which then yields (10). Thus, (11) is analytic by clause (ii) of our definition even though it is not a logical truth.

Nothing in this account entails anything about what constitutes the truth-value of an analytically

true statement. Nor does it entail any unrevisability claim. Thus, Quine's Belief-Revisability Argument does not apply to this form of analyticity, at least not in any obvious way. Again, following Boghossian (1997), I will call this general account of analyticity 'Frege-analyticity.'

Is there a connection between the *a priori* and Frege-analyticity? If *a priori* knowledge is genuine, then logical truths like (10) are *a priori* knowable. Having the appropriate sense experiences is necessary for acquiring the concept of dog, but beyond that no empirical evidence is needed to know that if something is a dog, then it is a dog. Grasping the conditions under which (10) is true suffices for knowing that it is true, regardless of whether there are dogs and regardless of whether one has observed any dog. (11) would also be *a priori* since mere rational reflection and understanding of what the two synonyms mean provides one with the sufficient warrant needed to be justified in believing (11), and hence in knowing (11) *a priori*. Generally speaking, if there are any *a priori* knowable truths at all, then all Frege-analytic truths are *a priori* knowable for the kinds of reasons just mentioned.

Not all *a priori* knowable truths are Frege-analytic, however. For example, we said earlier that (5a), (6a), and (7a) are *a priori* knowable:

- (5a) The relation *x is taller than y* is asymmetrical and transitive.
- (6a) Whatever is colored is extended.
- (7a) Whatever is red all over is not blue.

None of them is a Frege-analytic truth since none is a logical truth, nor can any of them be transformed into one by substitution of synonyms, *salva veritate*. If anything, (5a), (6a), and (7a) are examples of synthetic *a priori* truths, assuming that the distinction between Frege-analytic and Frege-synthetic truths is genuine, exhaustive, and mutually exclusive. This notion of analyticity won't be useful, then, to logical positivists who attempt to reductively explain all *a priori* truths in terms of analyticity.

There are Frege-analyticities if there are logical truths, which Quine admitted there are. That is important for two reasons. For one thing, as we saw, Quine thinks that any logical truth is revisable. If he is right, then some Frege-analytic truths are rationally revisable; in that case, his Quine's Belief-Revisability Argument won't apply to this kind of analytic truth. For another, Quine himself admits in Quine (1951) to having a clear idea of a logical truth whose meaning, but not its truth-value, is implicitly fixed by stipulative definition, even though he thinks that logical truths are open to possible rational

revision. It would be inaccurate to say, then, that Quine thought there were no analytic truths in any sense whatsoever. However, he did think that were sense in which there are analytic truths, namely, logical truths, is philosophically uninteresting and trivial, and is certainly not robust enough to erect a substantive epistemology of the sort that rationalists envision.

Quine also denies that there are Frege-analyticities of the second variety: sentences that can be transformed into logical truths by substitution of synonyms by synonyms. These are the ones that overlap with the conceptually necessary truths that Ayer and Carnap claimed we can know *a priori*, and quite cheaply, without sacrificing empiricism's main tenets about meaning or empirical content. Quine was dubious of conceptually necessary truths of any sort, so we should present his general argument for the denial of Frege-analytic truths of the second kind. (I count the following, along with Frege, as defenders of Fregean-analyticity, with their own nuanced views about synonymy and belief-revision: Chomsky, 1988, 2000; Katz, 1967, 1972, 1990; Putnam, 1962, 1965. Gilbert Harman and William Lycan are opponents of Frege-analyticity for largely Quinean reasons; see Harman, 1967a, 1967b; Lycan, 1994.)

Quine's Objections to Frege-Analyticity of the Second Kind

For Quine, the crucial question is whether we can explain synonymy without presupposing, either directly or indirectly, any notion of analyticity, given that we are trying to explain analyticity. In his 1951 study, Quine considers five analyses of synonymy and argues that they fail on independent grounds or they violate his noncircularity constraint.

The first proposal he considers was that two expressions are synonymous if and only if they have the same meaning. The problem is that the notion of two expressions having identical meanings is just as obscure and in need of explanation as the notion of analyticity. An account of synonymy that doesn't appeal to such mysterious entities as meanings is preferable to one that posits them.

The second proposal is that two expressions are synonymous just in case they can be mutually interchanged in any simple, **extensional**, sentence without changing its truth-value. This account fails to give a sufficient condition for synonymy. The predicates 'is a (normal healthy) creature with a heart' and 'is a (normal healthy) creature with a kidney' are true of exactly the same things. So, one could replace the former with the latter in sentence (12), resulting in sentence (13), without a change in truth-value:

- (12) Every creature with a heart is a creature with a kidney.
- (13) Every creature with a heart is a creature with a heart.

However, the two predicates are not synonymous, since (13) is supposedly analytic and *a priori* but (12) is neither.

The third proposal is that two expressions are synonymous just in case they can be interchanged in any **intensional** sentence without changing its truth-value. As Quine points out, the restrictive interchangeability criterion fails his first constraint. Consider sentences (14) and (15):

- (14) Necessarily, all and only brothers are brothers.
- (15) Necessarily, all and only brothers are male siblings.

According to the criterion, if 'brother' is synonymous with 'male sibling,' then (15) has to be true since (14) is. However, we can't determine whether (15) is true unless we first know whether the sentence 'all and only brothers are male siblings' is analytic. After all, the only reason one would have for thinking that (15) is true is if one believed that its embedded sentence expressed a conceptual truth. But to say that it is a conceptual truth that all and only brothers are male siblings is tantamount to saying that it is an analytic truth, which violates Quine's noncircularity constraint.

The final criterion for synonymy that Quine considered is this: two expressions are synonymous if they are interdefinable. Appealing to definitions is unhelpful since, as Quine noted, there are different kinds of definitions. For example, dictionary definitions are reports of prior synonymies, in which case the final criterion is circular, and thus trivial. Nor will it do to say, as Ayer and Carnap have, that a definition is simply a semantic rule or meaning postulate. By that understanding, two expressions are synonymous if and only if they are correlated with one another in virtue of some (non-empirical) semantic rule of the language. The label 'semantic rule' is arbitrary, implying only the appearance of a particular sentence (or a schema) under the heading 'Semantic Rules.' At best, a semantic rule of a language L can only specify which sentences (if any) of L have the honorific status of being analytic-in-L if there are any, but it won't help explain the absolute, nonrelativistic, sense of analyticity. Moreover, different sets of truths could have appeared under that heading without a change in the total set of truths in the language. Not only is the appeal to semantic rules and meaning postulates arbitrary and unmotivated, but also it cannot settle decisively the question of whether a

sentence like 'everything that is green is extended' is true by definition, or whether it expresses an important empirical truth about the world. Presumably it is one or the other, and it cannot be both. Since there are many borderline cases of just this sort, the Fregean analytic/synthetic distinction is unprincipled. On the assumption that his 'divide-and-conquer' strategy is successful, Quine concluded (Quine, 1951) that the notion of Frege-analyticity is empty and thus can serve no useful theoretical explanatory purpose.

Grice and Strawson (1956) is a famous response to Quine's argument; cf. Putnam (1965, 1970). According to Grice and Strawson, the terms 'analytic' and 'synthetic' must mark a genuine distinction, since they have a determinate philosophical use. Generations of philosophy students have learned these terms, and many of them can instantly recognize the real differences between the standard examples. Moreover, those who are adept in applying the distinction will agree about new cases. No doubt there will be difficult cases, as in 'everything that is green is extended' or perhaps 'all cats are animals,' but it does not follow that the analytic/synthetic distinction is not genuine. (Chomsky, 2000, makes a similar point.) The fact that there are borderline cases doesn't show that there are no clear cases of the distinction. Suppose there is no genuine difference in semantic status between analytic and synthetic sentences. Then there couldn't be any clear cases of analytic and synthetic truths or any agreement on the application of these terms to new cases. And yet, as they point out, even Quine would admit that a logical truth (e.g., 'all dogs are dogs') is analytic and not synthetic in some intuitive sense. According to Grice and Strawson, Quine's adequacy constraints on any plausible account of analyticity or synonymy are too stringent, unreasonable, and arbitrary; they have the undesired effect of rendering any linguistic elucidation of a group of family-related concepts impossible. In their view, Quine has only shown that no satisfactory analysis of analyticity has yet been provided, and not that no such distinction could be made. (Grice and Strawson never said what analyticity amounts to; nor did they show that the analytic/synthetic distinction is exhaustive.)

Until the late 1990s, much of the "analytic/synthetic" literature since the 1950s focused on the question of whether Quine succeeded in undermining the distinction. The prevailing view since Quine (1951) is that Quine did indeed win the battle, and that he also won the war against logical positivists, who tried to salvage empiricism while granting that we have knowledge of certain *a priori* necessary truths; see Soames (2003). Since the 1990s, there has been a resurgence of rationalism, and thus a reaction against

Quine's epistemology and his arguments against the tenability of any substantive analytic/synthetic distinction. The writings of Bealer, Boghossian, Bonjour, Katz, and Peacocke helped fuel this resurgence. I will conclude this article with Boghossian's main contribution to the analytic/synthetic debate.

Epistemological Analyticity

Paul Boghossian defends an epistemic notion of analyticity in the spirit, if not the letter, of Ayer (1946) (see Boghossian, 1996, 1997): *A sentence is analytic if and only if grasping what the sentence means suffices in one's being justified in believing in the truth of the proposition that the sentence expresses.* Ayer and Boghossian differ in their emphasis: Ayer explained the *a priori* in terms of a linguistic notion of analyticity, whereas Boghossian explains analyticity in terms of *a priori* justification. They also differ as to whether implicit definitions can literally determine the truth-value of a sentence: Ayer thought they can, but Boghossian does not. Boghossian's motivation for defending an epistemic account of analyticity was that he thought it had the best chance of explaining our having nonempirical knowledge of logical truths and logically valid inferences – assuming that the meaning of a logical constant is fixed by its conceptual role in formally valid patterns of logical inference, which may be something we can grasp by means of an implicit definition. (Boghossian thinks that, regardless of how one conceives of analyticity, one has no choice but to take analyticity as a genuine substantive property of certain sentences if one is a realist about meaning, which he is. For space constraints, I will sidestep the meaning/realism issue in this article.)

Suppose that you utter (16) and you go on to assert (17) (the example is from Lycan, 1994, although the main idea goes back to Ayer and Kripke):

(16) By 'veline' I shall henceforth mean 'a vegetarian cat.'

(17) All veline cats are vegetarians.

Then, just by virtue of grasping the content of your utterance of (16), which fixes the meaning of 'veline' in your idiolect, you are justified in believing that the proposition you expressed by your utterance of (17) is true. Hence, (17) (as tokened by you) is epistemically analytic.

Notice that, on Boghossian's epistemic account of analyticity, one can explain why such statements as (5a), (6a), (7a), (10), and (11), are *a priori*:

(5a) The relation *x* is taller than *y* is asymmetrical and transitive.

(6a) Whatever is colored is extended.

(7a) Whatever is red all over is not blue.

(10) Every brother is a brother.

(11) Every brother is a male sibling.

In each case, we have a *prima facie a priori* warrant to believe the proposition expressed because we have, in each case, a sufficient grasp of an implicit definition of an expression's meaning – either tacitly in our behavioral use of the expression or by way of an explicit stipulative definition. According to Boghossian, and apparently Kripke, our grasp of an implicit definition, which creates a semantic fact, constitutes our justification for believing the proposition expressed by the sentence containing the expression.

As Boghossian notes, Quine had no qualms about stipulative definitions if they are just arbitrary rules that implicitly define the meaning of a symbol. Thus, Quine could not object to the use of stipulative definitions *tout court*. He would object to the suggestion that stipulative definitions can literally create – by linguistic fiat or by linguistic convention – factual truths and valid inferences that are **unrevisable in principle**; see Quine (1935, 1960, 1966). However, Boghossian never claimed that implicit definitions determine the truth-value of a sentence or that they determine the validity of a logical inference. Rather, he claimed that implicit definitions literally create meaning-constituting facts the cognitive grasp of which sufficiently warrants semantic belief. If anything, they determine the proposition expressed but not the proposition's truth-value. Thus, as far as Boghossian is concerned, Quine's arguments (1935 and 1966) at best show that, contra Ayer and Carnap, implicit definitions cannot literally create truths by convention or by speaker intention. They don't show, argued Boghossian, that implicit definitions cannot have a meaning-constituting-factual role or an epistemic role in an explanation of *a priori* knowledge of logic.

Gilbert Harman, who defends the Quinean view, rejected Boghossian's epistemic account of analyticity and the role of implicit definitions in an explanation of *a priori* knowledge; see Harman (1996). According to Harman, grasping the contents of one's stipulative linguistic intentions can never warrant semantic belief. The reason is that meaning stipulations are just working assumptions that can easily be given up and, as such, they provide no evidence for anything. They can no more create epistemic warrant for semantic claims than they can create truth-values from scratch. See also Lycan (1994) for objections to Boghossian's view.

Readers can be assured that this debate will continue. We can look forward to more interesting discussions on the three distinctions discussed in this article.

See also: A Priori Knowledge: Linguistic Aspects; *De Dicto* versus *De Re*; Definitions: Uses and Varieties of; Empiricism; Holism, Semantic and Epistemic; Indeterminacy, Semantic; Meaning: Cognitive Dependency of Lexical Meaning; Meaning: Overview of Philosophical Theories; Possible Worlds: Philosophical Theories; Radical Interpretation, Translation and Interpretationalism; Realism and Antirealism; Rigid Designation.

Bibliography

- Ayer A J (1946). *Language, truth, and logic*. New York: Dover.
- Bealer G (1992). 'The incoherence of empiricism.' *The Aristotelian Society, Supplementary Volume* 66, 99–138.
- Bealer G (1996). 'A priori knowledge and the scope of philosophy.' *Philosophical Studies* 81, 121–142.
- Bealer G (1999). 'A theory of the a priori.' *Philosophical Perspectives* 13, 29–55.
- Boghossian P (1996). 'Analyticity reconsidered.' *Noûs* 30, 360–391.
- Boghossian P (1997). 'Analyticity.' In Hale B & Wright C (eds.) *A companion to the philosophy of language*. Oxford: Blackwell. 331–368.
- Boghossian P & Peacocke C (eds.) (2000). *New essays on the a priori*. Oxford: Oxford University Press.
- Bonjour L (1998). *In defense of pure reason*. Cambridge: Cambridge University Press.
- Carnap R (1956). *Meaning and necessity* (2nd edn.). Chicago: Chicago University Press.
- Casullo A (ed.) (1999). *A priori knowledge*. The International Research Library of Philosophy. Aldershot: Dartmouth Publishing Company.
- Casullo A (2003). *A priori justification*. New York: Oxford University Press.
- Chomsky N (1988). *Language and problems of knowledge*. Cambridge, MA: MIT Press.
- Chomsky N (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Coffa J (1991). *The semantic tradition from Kant to Frege*. Cambridge: Cambridge University Press.
- Devitt M (1994). *Coming to our senses*. Cambridge: Cambridge University Press.
- Driver J (2002). *Possible worlds*. London: Routledge.
- Duhem P (1914/1954). *The aim and structure of physical theory*. Weiner P (trans.). Princeton: Princeton University Press.
- Frege G (1968). *The foundations of arithmetic*. Austin J L (trans.). Oxford: Blackwell.
- Gendler T S & Hawthorne J (eds.) (2002). *Conceivability and possibility*. Oxford: Oxford University Press.
- Grice H P & Strawson P (1956). 'In defense of a dogma.' *Philosophical Review* LXV, 141–158. [Reprinted in Sleigh (1972) and in Sumner & Woods (1969).]
- Harman G (1967a). 'Quine on meaning and existence, I: the death of meaning.' *Review of Metaphysics* 21, 124–151.
- Harman G (1967b). 'Quine on meaning and existence, II: existential commitment.' *Review of Metaphysics* 21, 343–367.
- Harman G (1996). 'Analyticity regained?' *Noûs* 30, 392–400.
- Kant I (1781/1998). *The critique of pure reason*. Guyer P & Wood A W (trans.). Cambridge: University of Cambridge Press.
- Katz J J C (1967). 'Some remarks on Quine on analyticity.' *The Journal of Philosophy* LXIV, 36–52. [Reprinted in Sumner & Woods (1969).]
- Katz J J C (1972). *Semantic theory*. New York: Harper and Row.
- Katz J J C (1990). *The metaphysics of meaning*. Cambridge, MA: MIT Press.
- Katz J J C (2000). *Realistic rationalism*. Cambridge, MA: MIT Press.
- Kripke S (1972/1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Linsky L (ed.) (1977). *Reference and modality*. Oxford: Oxford University Press.
- Loux M (ed.) (1979). *The possible and the actual: readings in the metaphysics of modality*. Ithaca: Cornell University Press.
- Lycan W G (1994). *Modality and meaning*. Dordrecht: Kluwer Academic Publishers.
- Melia J M (2003). *Modality*. Montreal & Kingston: McGill-Queen's University Press.
- Moser P (ed.) (1987). *A priori knowledge*. Oxford: Oxford University Press.
- Pap A (1958). *Semantics and necessary truth*. New Haven: Yale University Press.
- Peacocke C (1999). *Being known*. Oxford: Clarendon Press.
- Peacocke C (2004). *The realm of reason*. Oxford: Oxford University Press.
- Putnam H (1962). 'It ain't necessarily so.' *The Journal of Philosophy* LIX, 658–671. [Reprinted in Putnam H (1979). *Philosophical papers* (vol. 1). Cambridge: Cambridge University Press.]
- Putnam H (1965). 'The analytic and the synthetic.' In Putnam H (ed.) *Philosophical papers*, vol. 2. Cambridge: Cambridge University Press, 1975.
- Putnam H (1970). 'Is semantics possible?' *Metaphilosophy* 1, 189–201. [Reprinted in Putnam H (1985). *Philosophical Papers* (vol. 2). Cambridge: Cambridge University Press. 139–152.]
- Putnam H (1983). '"Two Dogmas" revisited.' In Putnam H (ed.) *Philosophical papers*, vol. 3. Cambridge: Cambridge University Press, 1985. 87–97.
- Quine W V O (1935). 'Truth by convention.' In Quine W V O (ed.) *Ways of Paradox and Other Essays*, 2nd edn. Cambridge, MA: Harvard University Press, 1976. 77–106.
- Quine W V O (1951). 'Two dogmas of empiricism.' In Quine W V O (ed.) *From a logical point of view*. Cambridge, MA: Harvard University Press, 1980. 107–132. [Reprinted in Sleigh (1972) and in Sumner & Woods (1969).]

- Quine W V O (1960). *Word and object*. Cambridge, MA: MIT Press.
- Quine W V O (1966). 'Carnap and logical truth.' In Quine W V O (ed.) *Ways of Paradox and Other Essays*, 2nd edn. Cambridge, MA: Harvard University Press, 1976. 20–46.
- Quine W V O (1967). 'On a suggestion of Katz.' *The Journal of Philosophy* 64, 52–54. [Reprinted in Sumner and Woods (1969).]
- Quine W V O (1991). 'Two dogmas in retrospect.' *Canadian Journal of Philosophy* 21, 265–274.
- Salmon N (1993). 'Relative and absolute apriority.' *Philosophical Studies* 69, 83–100.
- Sleigh R C (ed.) (1972). *Necessary truth*. Englewood Cliffs: Prentice-Hall.
- Soames S (2003). *Philosophical analysis in the twentieth century: the age of meaning*, vol. II. Princeton: Princeton University Press.
- Sumner L W & Woods J (eds.) (1969). *Necessary truth: a book of readings*. New York: Random House.

Anaphora: Philosophical Aspects

J King, University of Southern California, Los Angeles, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

Anaphora is the phenomenon whereby an expression, which is called a **proform** (e.g., 'he' [pronoun], 'so' [proadjective]), is interpreted in light of another expression in its immediate linguistic context, which is called the proform's **antecedent**. In this article, I will confine my attention to anaphoric pronouns, since they are the instances of anaphora that are of most interest to philosophers and linguists. A simple example of an anaphoric pronoun is the following (where 'Rebecca' is the antecedent of 'she'):

- (1) Rebecca left and she was angry.

Simple instances of anaphora such as (1), in which an anaphoric pronoun has a referring expression as its antecedent, are not of much interest to philosophers and linguists. The reason is that such cases seem easy to understand. The anaphoric pronoun simply refers to whatever its antecedent does.

Another sort of case in which the behavior of an anaphoric pronoun is well understood, and hence not of much current interest to philosophers and linguists, is a case of the following sort (on the reading on which 'she' has 'Every professor' as its antecedent):

- (2) Every professor believes she is smart.

In such cases, it is widely held that the pronoun is functioning as a variable bound by its quantifier antecedent. Thus, (2) could be represented more perspicuously as:

- (2a) [Every professor: x][x believes x is smart]

where the third occurrence of 'x' corresponds to the pronoun. Since the pronoun behaves like a variable in predicate logic, and since the behavior of such variables is well understood, most philosophers and linguists do not worry about such cases.

The interest of linguists and philosophers in anaphora derives from the fact that there are instances of anaphora in which the anaphoric expression cannot be understood as an expression referring to whatever its antecedent does (as in [1]) nor as a variable bound by its quantifier antecedent (as in [2]). Let's call such cases instances of **puzzling anaphora**. The question then arises as to how puzzling anaphora should be understood semantically. As will be seen, there are a number of competing proposals. We shall discuss two kinds of puzzling anaphora here. The first instance of puzzling anaphora is a case in which a pronoun has a quantifier antecedent in another sentence. Anaphora of this sort is often called **discourse anaphora**. Here are two examples:

- (3) Few students passed the exam. They studied hard.
- (4) A woman is following Glenn. She is from the IRS.

For the moment, let's concentrate on (3). There are at least two reasons for thinking that the pronoun in (3) is not a variable bound by its quantifier antecedent. Both reasons are discussed by Evans (1977). The first is that such a treatment clearly yields the wrong truth conditions for examples like (3). If 'they' is a bound variable in (3), the two sentences of (3) together should be equivalent to:

- (3a) Few students: x (x passed the exam and x studied hard)

But this is obviously incorrect, since (3) entails that few students passed the exam, but (3a) does not (it would be true if many students passed, but few both passed and studied hard).

Second, it is generally thought that the scope of a quantifier cannot extend beyond the sentence in which it occurs. If that is correct, then the pronoun in (3) falls outside of the scope of its quantifier antecedent and so cannot be bound by it.

Further, the pronoun in (3) cannot be understood as referring to whatever its antecedent does, since its antecedent is a quantifier and not a referring term at all. Since the pronoun cannot be understood as a bound variable nor as an expression that refers to whatever its antecedent does, the question arises as to how we should understand it semantically.

A second kind of puzzling anaphora is called **donkey anaphora** and it comes in two varieties. The first I will call **conditional donkey anaphora**:

(5) If Sarah owns a donkey, she beats it.

(where 'it' is anaphoric on 'a donkey'). The second I will call **relative clause donkey anaphora**:

(6) Every woman who owns a donkey beats it.

(where 'it' is anaphoric on 'a donkey'). On the readings we are concerned with, neither (5) nor (6) is talking about any particular donkey, and so the pronoun 'it' cannot be a term referring to a particular donkey. Further, in the case of (5), all independent evidence available suggests that a quantifier cannot take wide scope over a conditional and bind variables in its consequent (*'If John owns every donkey_i, he beats it_i'). This suggests that the (apparent) quantifier 'a donkey' in (5) cannot bind the pronoun in the consequent. In addition, even if 'a donkey' could magically do this in (5), assuming it is an existential quantifier, we still would not get the intuitive truth conditions of (5), which require that Sarah beat every donkey she owns. Similarly, the independent evidence available suggests that quantifiers cannot scope out of relative clauses (*'A man who owns every donkey_i beats it_i'), and so again the pronoun in (6) is not within the scope of its quantifier antecedent and so is not bound by it. Thus, the pronouns in both conditional and relative clause donkey anaphora cannot be understood as referring expressions nor as bound variables.

So we have two kinds of puzzling anaphora, discourse anaphora and donkey anaphora. In neither case can the anaphoric pronouns be understood as bound variables or as expressions referring to whatever their antecedents refer to. The question is: how should we understand such cases of anaphora semantically? A number of theories have been developed to answer this question. Sketches of four prominent theories of puzzling anaphora are given here. Because the theories are technical and sophisticated, only the barest sketch can be given of them. Readers interested in more detail should consult papers cited in the bibliography.

The first theory of puzzling anaphora we will consider is **Discourse Representation Theory** or **DRT**. Though this name for the view comes from Kamp (1981), Heim (1982) independently developed a

very similar view. Here I will be neutral between the somewhat different Heim and Kamp formulations. DRT claims that the indefinites in discourse anaphora like (4) ('A woman') and donkey anaphora like (5) and (6) ('a donkey') are not quantifiers. Instead they are really predicates that introduce free variables. The anaphoric pronouns are variables in such cases too. The apparent existential force of the indefinite in (4) results from default existential quantification over free variables. In (5), the semantics of the pronoun interacts with the semantics of the conditional to yield the result that the sentence is true iff Sarah beats every donkey she owns. In (6), the quantifier 'Every woman' 'unselectively binds' both the woman variables in logical form and the variables introduced by 'a donkey' and 'it'. So in (6), the anaphoric pronoun is bound after all, but not by its antecedent ('a donkey').

Descriptive Approaches to anaphora have been formulated by Evans (1977), Parsons (1978), Davies (1981), Neale (1990), and Heim (1990). On approaches of this sort, anaphoric pronouns in puzzling anaphora in some sense are interpreted as definite descriptions. Thus, the pronoun in (4) is interpreted as 'the woman following Glenn'. Similar remarks apply to the pronouns in (5) and (6). Here different accounts are given as to why and how (5) and (6) are true iff Sarah/Every woman beats every donkey she owns. For example, Neale (1990) claims that in such cases, the pronouns are interpreted as 'numberless descriptions,' which are essentially universal quantifiers. This yields the truth conditions mentioned. For a different approach, see Heim (1990).

The Context-Dependent Quantifier Approach (CDQ) is formulated and defended by King (1987, 1991, 1994, 2005) and Wilson (1984). On this approach, pronouns in instances of puzzling anaphora are themselves quantifiers. The forces (universal or existential), restrictions (what the quantifiers range over – e.g., 'every student' ranges over students) and relative scopes are determined by features of their linguistic contexts. Thus, according to CDQ, the anaphoric pronouns in puzzling anaphora are contextually sensitive devices of quantification. The precise natures of the quantifiers they express are determined by features of their linguistic contexts. In (4), CDQ holds that the pronoun expresses the existential quantifier normally expressed outside of any context by the indefinite 'a woman following Glenn'. Similar remarks apply to (5) and (6), except that the semantics of the conditional interacts with the pronoun *qua* quantifier to get the proper reading of (5) (see King, 2005); and in (6), the pronoun *qua* quantifier takes narrow scope with respect to the universal quantifier 'Every woman' (because its antecedent does as well).

Finally, there are **Dynamic Logic Accounts (DL)** of puzzling anaphora, originally formulated by Groenendijk and Stokhof, (1991), with different versions put forward by Chierchia (1995) and Kanazawa (1994a and 1994b). The key claim of such accounts is that quantifiers can semantically bind pronouns even if those pronouns do not occur in their syntactic scopes. So, despite appearances, the pronouns in (4), (5), and (6) are semantically bound by their quantifier antecedents. DL also makes novel claims about the semantics of conditionals and universal quantification to get the truth conditions of (5) and (6) to come out right.

At this stage of inquiry, it is fair to say that there is no consensus on which of the above-mentioned theories is correct. Each of the theories faces some difficulties and it is difficult to determine which theory is most successful. It is to be hoped that future research will shed more light on the proper theory of puzzling anaphora.

See also: Deixis and Anaphora: Pragmatic Approaches.

Bibliography

- Chierchia G (1995). *Dynamics of meaning*. Chicago and London: The University of Chicago Press.
 Davies M (1981). *Meaning, quantification and necessity*. London: Routledge and Kegan Paul.

- Evans G (1977). 'Pronouns, quantifiers and relative clauses (I).' *Canadian Journal of Philosophy* VIII(3), 467–536.
 Groenendijk J & Stokhof M (1991). 'Dynamic predicate logic.' *Linguistics and Philosophy* 14, 39–100.
 Heim I (1990). 'E-type pronouns and donkey anaphora.' *Linguistics and Philosophy* 13, 137–177.
 Heim I (1982). *The semantics of definite and indefinite noun phrases*. Ph.D. diss. University of Massachusetts.
 Kamp H (1981). 'A theory of truth and semantic representation.' In Groenendijk J & Stokhof M (eds.) *Formal methods in the study of language*. Amsterdam: Mathematical Centre.
 Kanazawa M (1994a). 'Dynamic generalized quantifiers and monotonicity.' *Dynamics, Polarity and Quantification*. CSLI Lecture Notes No. 48. Stanford: Center for the Study of Language and Information.
 Kanazawa M (1994b). 'Weak vs. strong readings of donkey sentences in a dynamic setting.' *Linguistics and Philosophy* 17(2), 109–158.
 King J C (1987). 'Pronouns, descriptions and the semantics of discourse.' *Philosophical Studies* 51, 341–363.
 King J C (1991). 'Instantial terms, anaphora and arbitrary objects.' *Philosophical Studies* 61, 239–265.
 King J C (1994). 'Anaphora and operators.' *Philosophical Perspectives* 8, *Logic and Language* 8, 221–250.
 King J C (2005). 'Context dependent quantifiers and donkey anaphora.' *Canadian Journal of Philosophy* (forthcoming).
 Neale S (1990). *Descriptions*. Cambridge: MIT Press.
 Wilson G (1984). 'Pronouns and pronominal descriptions: a new semantical category.' *Philosophical Studies* 45, 1–30.

Architecture of Grammar

R Kempson, King's College London,
 London, UK

© 2006 Elsevier Ltd. All rights reserved.

A grammar of a language is by definition a formal model of the properties of language that are intrinsic to the words of the language and the way they can be combined. There are then a number of choices to make:

1. Should the explanation be defined solely to reflect structural properties of language, or should those properties in some sense be correlated with the way language is used in speaking or hearing?

This is the issue of language competence and its separation from language performance.

There is another criterion which a grammar might be required to meet, which might seem like just another way of putting the same question, but is taken as requiring a different answer:

2. Is it enough to provide a formal account of language structure, or should a grammar reflect general cognitive underpinnings, with the theory of grammar seen as part of a theory of mind?

This is the question of language as formal system vs. language as part of cognitive psychology.

There are then two debates about the internal design of the grammar to be considered. The first concerns the status of semantics in the grammar; the second concerns the status of the lexicon and its relation to syntax:

3. Should the grammar define a direct relation between the expressions of the language, and the entities in the world they are used to talk about; or should interpretation for natural-language expressions be defined in terms of concepts that the words express?

4. Is there some central body of properties that constitute the syntactic system of language, to be defined independently of the words that make up

the language, or is the entire body of structural properties of language expressible as statements about words and their combinatorial properties?

These issues are not unrelated, as will emerge; but the following discussion takes these issues in turn, keeping them separate in so far as possible.

The Competence-Performance Divide

Even though interpretation of utterances invariably depends on the context in which that sentence is uttered, in seeking to explain natural language, linguists generally restrict their attention to those properties of language that are intrinsic to the words of the language and their mode of combination, defining grammars for individual languages to capture those aspects of sound-interpretation correlation that are not subject to variation context by context. This separation of linguistic analysis from any direct reflection of language as used in context led to what are called competence models of language (Chomsky, 1965, 1986, 1995). These define principles that determine the limits on sound-meaning correspondences both for any one language and for language in general; and they are articulated without consideration of any application in language use. Models of performance are then defined to articulate the mechanisms necessary to use such models in parsing and production (Fodor, 1981; Garrett, 2000; Gorrell, 1995; Levelt, 1989) and are taken to be dependent on the prior and independent articulation of such competence models. There is an immediate consequence to this separation. In virtue of the lack of direct correspondence between competence models and data culled from language performance, grammars can only be evaluated relative to a relatively abstract concept of data; and these data are judgements of wellformedness of sentence plus interpretation pairs as made by competent speakers of the language.

Grammars: Formal-Language Theory vs. Psychology

Reflecting such string-interpretation judgements, the architecture defined is modular, with more than one level of representation defined, in order to capture the distinct forms of generalization needed. Minimally, there are three kinds of generalization – phonological, syntactic and semantic – reflecting the existence of mappings from phonological form onto interpretation via some projection of intermediate structure, with the lexicon listing words with their phonological, syntactic, and semantic specifications. (Since not all languages are written, spoken language is taken

to be primary.) The number and nature of such levels and the relation between them is, however, subject to debate. The debate is founded in two distinct points of departure for defining a grammar formalism, two different ways, that is, of answering question (2), which require two different ways of answering question (3).

The Formal-Language Pattern

On the one hand lies the formal-language tradition in which an articulated purpose-built formal system devised for the formal study of some (natural) phenomenon has a defined syntax, with an associated mapping from each licensed string onto some subset of the phenomenon under analysis. In such systems, the commitment to defining a concept of denotational content for the expressions of the language is definitional: the expressions pick out the objects under analysis, their denotation, for that is what the formal language is designed to do. The concept of compositionality is also true by definition: the meaning of a compound expression of the language is exclusively determined by the meaning of the parts and their mode of combination. Natural languages have been treated in like manner, explicitly following the formal language pattern (Montague, 1974). Categorical grammars, in particular, are faithful advocates of this formal-language tradition. They add to the characterization of a grammar some presumed specification of phonological properties of the listed primitive terms, but otherwise preserve the general architecture of a pairing of strings and denotational contents, with syntax no more than a systematic procedure for constructing pairings of phonological complexes and denotational contents for progressively larger units from words to sentences, with syntax and semantics essentially in tandem (Moortgat, 1997; Morrill, 1994; Ranta, 1994; Steedman, 1996, 2000). As might be expected given the adherence to the formal-language pattern, comprehensive accounts of anaphora and attendant concepts of underspecification of denotational content are commonly not provided (though see Ranta, 1994 for a proof-theoretic reconstruction of context and anaphoric dependency), and advocates of categorical grammar formalisms generally do not evaluate their formalisms relative to criteria of psychological reality (though see Morrill, 2002 for exploration of using categorical-grammar constructs in the study of acceptability). Consequently, the categorical grammarian's answer to question (2) is to reject the need for cognitive underpinnings. And the answer to question (3) is that a direct mapping from natural-language expression to denotational content is central to natural-language grammar formalism.

The Language Faculty as a Subpart of Central Cognitive Systems

On the other hand, there is the psychology-based tradition, committed to the stronger form of answer to (2), and denying the formal-language stance on (3). Within the general cognitive-science perspective, processing of input stimuli by a cognitive system (seeing, hearing, smelling, etc.) involve special-purpose devices that enable internal representations to be constructed on the basis of which such input stimuli are assigned interpretation by the cognitive system, the human agent then using these internally constructed representations to reason to conclusions about what it is that he has seen/heard/smelled, etc. Minimally, on this view, modeling operations of the cognitive system involves defining a mapping from representations of those stimuli onto representations that are formulae of a deductive logical system (or systems), the internalized “language(s) of thought” (Fodor, 1983). The representation thus constructed as interpretation of the input stimulus may have to be assigned some denotational content underpinning the constructed interpretation depending on one’s theory of mind, but it is the representation itself that constitutes the interpretation of the processed stimulus. There is a second property of such input systems, which natural language illustrates well. In all cases of input to the cognitive system, there is a gap between the information derived directly from some low-level representation of the input stimulus and the interpretation assumed by the cognitive system to be derived from it. This gap, which is essential to the flexibility and finite resources of the cognitive system itself, is filled by some mode of so-called non-demonstrative reasoning using other information as made available in the context. This perspective on natural-language modeling provides a very different starting point from the formal-language tradition, presuming as it does on a mapping from one form of representation to another, of which the input forms are systematically less rich than the output forms. In natural languages, this is displayed by pronouns and elliptical constructions, which provide an input to the interpretation process that is systematically less specified than how they are understood on any occasion of their use, having to be enriched by information that context makes available, in order to establish what information the anaphoric/elliptical expression conveys.

Chomskian grammars are the major representative of grammars adopting the psychological tradition of representationalism broadly construed, with all concepts of interpretation for natural language string articulated only as logical form representations (at the defined level of LF, see Chomsky, 1986, 1995

and elsewhere). On this view, syntax is defined as the core computational system with the levels of phonological form (PF) and logical form (LF) constituting interfaces with external systems (phonetic and intensional/conceptual systems, respectively). Central to this system is the characterization of discontinuities observable in all natural languages in which some expression may be indefinitely far from some position in the string with which, for interpretation purposes, it needs to be paired. These are cited by Chomsky as evidence that natural languages are not perfect formal systems (Chomsky, 1995), this departure justifying the irreducibility in the grammar architecture of the computational (syntactic) system. In capturing the supposed independence of the language computational system from the external systems with which it is interfaced, the minimalist system nevertheless respects general cognitive assumptions in so far as the internal properties of the language module have to be defined to allow interaction with general cognitive processes through these interface levels. As we would expect, given these assumptions, the properties of anaphora and ellipsis construal that are articulated as within the grammar formalism are expressed as syntactic principles.

Combining Formal Language and Psychological Traditions

Outside the Chomskian tradition, the three major grammar formalisms, Head-Driven Phrase-Structure Grammars (HPSG; Sag *et al.*, 2003), Tree-Adjoining Grammars (TAG; Joshi, 1987; Joshi and Kulick, 1995) and Lexical Functional Grammar (LFG; Bresnan, 2001; Dalrymple, 2001), all adopt a mixed approach, articulating semantics in terms of denotational content, but nevertheless preserving some form of cognitivist perspective.

On the one hand, they reflect the formal-language tradition in assuming that a specification of denotational semantics for sentence strings is part of the grammar formalism. Even while respecting this assumption, they nevertheless depart from the restrictiveness of the syntax-semantics correspondence defined for formal languages. Building on the presumed independence of phonological generalizations from the remainder, they allow the independence of syntactic generalizations from the combinatorial system required for other levels. Grammar formalisms may be defined to allow an *n*-tuple of levels, each articulated with independent vocabulary and forms of generalization (see Dalrymple, 2001; Jackendoff, 2002), leading to the consequent challenge of articulating adequately restrictive correspondence rules between these independent levels (Dalrymple, 1999). Despite the

relaxation of the tight architecture of formal-language grammar formalisms, the assumptions of compositionality, broadly construed, are preserved, with each word being assigned some n-tuple of phonological, syntactic, semantic, morpho-syntactic, information-structure, etc. each of which is in the form of constraints on its combinatorial potential, which then unify with specifications assigned to its neighboring items according as their specifications are consistent. Hence the term “unification-based grammars.”

Notwithstanding their conforming broadly to formal-language assumptions, such grammars lay claim to assumptions of psychological reality not dissimilar to those laid out by Chomsky. In so far as they invoke a level of denotational semantics as part of the grammar formalism, however, they fail to be commensurate with the representationalist methodology. And some of them blur the competence/performance distinction in allowing a pragmatic component within the grammar formalism (Sag and Ginzburg, 2003), a problematic move, since pragmatics is the study of the general cognitive principles underpinning successful communication, hence not part of a natural-language grammar.

Issues of Grammar Design

The Status of the Lexicon

The different status of the lexicon in the overall system reflects the different importance assigned to the lexicon in the two traditions. Following the formal language tradition, categorial grammars, HPSG, and LFG are fully lexicalized, with the lexicon providing a list of input triples of phonological, syntactic and semantic specifications for each lexical entry, these constraint sets dictating the mode of combination of adjacent or otherwise co-occurring expressions in the phonological, syntactic and semantic domains, respectively. Earlier Chomskian paradigms assigned a smaller role to the lexicon, with syntactic principles determining the core computational actions that induce structure, hence in principle reflecting the psychological distinction between stored information to be retrieved from memory vs. on-line construction. Phonological and semantic interpretation processes were merely interpretive, interface, mechanisms applying to the two distinct types of output from the central syntactic device. The central status of syntax led to the Chomskian Principles and Parameters model of language being described as the T-model:

phonology ← syntax → logical form

Between the two extremes of full lexicalist and nonlexicalist approaches, TAG formalisms project,

for each word (lexical item) entered in the lexicon, a family of local structures is defined as a stipulated form of input, but the nesting of these into larger structures is the result of general tree-adjoining mechanisms. In minimalist approaches, the lexicon plays a greater role than in earlier Chomskian models, but nevertheless detailed lexical specifications are not formally articulated.

The Syntax–Semantics Debate and the Grammar–Parser Relation

The discussion so far has presumed on the distinctness of syntax and semantics. Upon the formal-language stance, these are by definition distinct, for one is a representation, the other, being denotational, is purely interpretive of such representations. Given the representationalist methodology, however, the question arises whether representations defined as the syntactic mechanism for natural-language grammars constitutes anything more than a mechanism for progressively building the conceptual representations that constitute an interpretation of the signal. With a shift of perspective to one in which grammar formalisms are defined to model the mapping from phonological signals onto a progressively richer conceptual representation following the time-linear dynamics of on-line processing (Kempson *et al.*, 2001; Phillips, 2003; Cann *et al.*, 2005), exactly this simpler stance becomes available, with the bonus of a much more direct reflection of language in use. What have been thought of as irreducibly syntactic generalizations, in particular the discontinuity effects made famous in Chomsky’s work as motivation for movement processes in syntax, become expressible as part of the progressive time-linear construction of the emergent conceptual representation. On this view, the architecture of the grammar formalism acquires a directional dynamics, with phonological specifications providing input to actions that induce the building up of conceptual representations, as determined jointly by contextually provided structure, lexical specifications, and general structure-building processes:

phonology → syntax → logical form

The syntactic mechanism becomes simpler because instead of inducing a distinct syntactic structure and articulating a mapping between syntactic and phonological representations on the one hand and between syntactic and semantic (i.e., conceptual) representations on the other (as in the T-model), there is only the progressive construction of conceptual representations from phonological input. But such a shift leads to a radical narrowing of the traditional competence-performance dichotomy, for the competence model

reflects the dynamics needed to model performance, and becomes directly embeddable into psycholinguistic models of the general cognitive system.

Summary

As these snapshots indicate, the architecture of the language faculty as defined in the various grammar formalisms reflect different claims about the status of language within an overall theory of mind. The significance of a parsing-directed design for natural-language grammars, introduced as a recent departure from orthodox assumptions, is that it leads to re-evaluation of the role of language in the theory of mind: any property of the architecture intrinsic to all natural language grammars is by assumption not merely a design feature of language itself, but a sub-part of a general theory of cognitive psychology and a reflection of the human capacity for language use.

Bibliography

- Bresnan J (2001). *Lexical functional syntax*. Oxford: Blackwell.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky N (1986). *Knowledge of language*. New York: Praeger.
- Chomsky N (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Cann R, Kempson R & Marten L (2005). *The dynamics of language*. Oxford: Elsevier.
- Dalrymple M (ed.) (1999). *Syntax and semantics in lexical functional grammar: the resource logic approach*. Cambridge, MA: MIT Press.
- Dalrymple M (2001). *Lexical functional grammar*. New York: Academic Press.
- Fodor J A (1981). *Modularity of mind*. Cambridge: MIT Press.
- Garrett M (2000). 'Remarks on the architecture of language processing systems.' In Grodzinsky Y, Shapiro L & Swinney D (eds.) *Language and the brain*. New York: Academic Press. 35–64.
- Gorrell P (1995). *Syntax and parsing*. Cambridge: Cambridge University Press.
- Jackendoff R (2002). *Foundations of language*. Oxford: Oxford University Press.
- Joshi A (1987). 'An introduction to Tree-Adjoining Grammars.' In Manaster-Ramer A (ed.) *Mathematics of language*. Amsterdam: John Benjamins. 87–114.
- Joshi A & Kulick S (1997). 'Partial proof trees as building blocks for a categorial grammar.' *Linguistics and Philosophy* 20, 637–667.
- Kempson R, Meyer-Viol & Gabbay D (2001). *Dynamic syntax*. Oxford: Blackwell.
- Levelt P (1989). *Speaking*. Cambridge, MA: MIT Press.
- Montague R (1974). Thomasson R (ed.). *Formal approaches to language*. Yale: Yale University Press.
- Moortgat M (1997). 'Categorial type logics.' In van Benthem J & ter Meulen A (eds.) *Handbook of logic and language*. Cambridge, MA: MIT Press. 93–179.
- Morrill G (1994). *Type-logical grammar*. Dordrecht: Foris.
- Morrill G (2002). 'Incremental processing and acceptability.' *Computational Linguistics* 26, 319–338.
- Phillips C (2003). 'Linear order and constituency.' *Linguistic Inquiry* 34, 37–90.
- Ranta A (1994). *Type-theoretic grammar*. Oxford: Clarendon Press.
- Sag I & Ginzburg J (2002). *Syntax and semantics of questions in English*. Stanford: CSLI.
- Sag I, Wasow T & Bender E (2003). *Formal syntax*. Stanford: CSLI Press.
- Steedman M (1996). *Surface structure and interpretation*. Cambridge, MA: MIT Press.
- Steedman M (2000). *The syntactic process*. Cambridge, MA: MIT Press.

Aristotle and Linguistics

P A M Seuren, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

The study of language has always had two kinds of practitioners, the practical and the theoretical linguists. Aristotle was no doubt the first theoretical linguist (in addition to being the first in many other subjects), but he also contributed essentially to the development of practical linguistics. His role in the history of linguistics has been highlighted in a few publications (e.g., Seuren, 1998; Allan, 2004).

Aristotle was born in Stagira, in Ancient Macedonia, in 384 B.C.E. His father was the personal physician and a close friend of the king of Macedonia, Amyntas II. An exceptionally gifted boy to begin with, Aristotle joined Plato's Academy in Athens at the age of 17, to remain there until Plato's death in 347. Having been passed over as Plato's successor, he left Athens to live, first, in Asia Minor and then in Lesbos. In 343–342, Amyntas' son and successor, Philip II of Macedonia, invited him to come and teach his son Alexander, then 14 years old. This he did for 2 years. In 336, Alexander succeeded his father and immediately conquered the whole of Greece. Under

Alexander's political protection, Aristotle returned to Athens in 335 and founded his school of philosophy, the Lyceum. There he taught until 323, when news of Alexander's death reached Athens. No longer certain of Macedonian protection, he left Athens overnight and sought refuge in Chalcis, just north of Athens, where a Macedonian garrison was stationed. One year later, in 322, he died of an intestinal disease.

His first great contribution to the study of language—not often mentioned—is the fact that he demythologized language. Rather than seeing language as a magical instrument to cast spells, entrance people, and call up past, present, and future spirits, he saw language as an object of rational inquiry, a means of expressing and communicating thoughts about anything in the world. The 'semiotic triangle' of (a) language as the expression of (b) thoughts that are intentionally related with (c) elements in the world, famously depicted in Ogden and Richards (1923: 11), is due to Aristotle. This is Aristotle's most general and perhaps also his most important contribution to the study of language, even if it is not often mentioned by modern authors, for whom it has become a matter of course that language can be seen as a valid object of rational inquiry.

In a more analytical sense, Aristotle's role in the development of linguistics is in large part due to his theory of truth. For him, truth and falsity are properties of either thoughts or sentences. A classic statement is (*Metaphysics* 1027b25):

For falsity and truth are not properties of actual things in the world (so that, for example, good things could be called true and bad things false), but properties of thought.

A few pages earlier, he defines truth as follows (*Metaphysics* 1011b26):

We begin by defining truth and falsehood. Falsehood consists in saying of that which is that it is not, or of that which is not that it is. Truth consists in saying of that which is that it is, or of that which is not that it is not.

Here Aristotle introduces *not* as a simple truth-functional inverter of truth values: a toggle between true and false. This has momentous consequences.

Aristotle's truth theory is known as the correspondence theory of truth, in that it requires a correspondence between what is the case in the world on the one hand and what is said or thought on the other. To make this notion of correspondence more explicit, some form of analysis is needed. Aristotle made a beginning with that. He analyzes the 'world' as consisting of things that are named by any of the 10 categories substance, quantity, quality, relation, place, time, position, state, action, or affection (*Categories* 1b25–30). Within the category 'substance,'

there is a hierarchy from the primary substances (individual existing entities) through a range of secondary substances, from species and genus to any higher order. The secondary substances together with the remaining 9 categories are properties or things that things are ("everything except primary substances is either predicable of a primary substance or present in it"; *Categories* 2a33).

On the other hand, he analyzes sentences as resulting from the application of a *katêgoroumenon* (Latin *praedicatum*) to something. The something to which the predicate is applied he calls *hypokeímenon* (literally 'that which underlies'; Latin *subiectum* or *suppositum*). Primary substances (entities) can be the object only of predicate application – that is, can only be *hypokeímena* (*Categories* 2b39–40). All other things can be either *hypokeímena* or properties, denoted by a predicate. Yet in orderly talk about the universe, it is proper to take lower categories of substance as the things predicates apply to and reserve predicates themselves for the denoting of higher-order substances and other categories of being (*Categories* 3a1–5).

The combination of a predicate with a term denoting the *hypokeímenon* Aristotle calls *prótasis* (Latin *propositio*). A proposition is true just in case the property assigned to the *hypokeímenon* actually adheres to it; otherwise it is false. Moreover, a true proposition is made false, and vice versa, by the prefixing of *not* ("it is not the case that"). The term *prótasis* occurs for the first time on the first page of *Prior Analytics*, which contains his doctrine of syllogisms (*Prior Analytics* 24a16):

A proposition (*prótasis*) is an affirmative or negative expression that says something of something.

A proposition is divided into terms (*Prior Analytics* 24b16):

A term (*hóron*) I call that into which a proposition is analyzed, such as the predicate (*katêgoroumenon*) and that to which the predicate is applied.

One notes that Aristotle lacked a word for what we call the subject term of a sentence. During the late Middle Ages, the Latin *subiectum* began to be used in that sense—an innovation that has persisted until the present time (Seuren, 1998: 121–124).

This was the first semantic analysis of sentence structure in history, presaged by, and probably unthinkable without, Plato's incipient analysis of sentence meaning in his dialogue *The Sophist*. It is important to note that Aristotle's analysis of the proposition does not correspond to the modern syntactic analysis in terms of subject and predicate, but rather to what is known as topic-comment analysis. The identification of Aristotle's sentence constituent for the denoting of a *hypokeímenon* with "grammatical

subject,” characterized by nominative case, and of Aristotle’s predicate with “grammatical predicate,” may have been suggested by Aristotle, as when he says that a morphological verb “always is a sign of something said of something else” (*On Interpretation* 16b7). But it was carried through systematically a few decades after Aristotle’s death by the linguists of Alexandria, whose task it was to develop teaching material for the Egyptian schools where local children had to learn Greek in the shortest possible time (Seuren, 1998: 21–22). Unfortunately, this identification was, though convenient, rash and ill-considered. It persisted more or less unchallenged until the middle of the 19th century, when some, mostly German, scholars discovered that the Aristotelian subject–predicate distinction does not coincide with the syntactic subject–predicate analysis universally applied in linguistics. For in actual discourse, very often what should be the subject according to Aristotle’s definition is not the subject recognized in grammatical analysis, and likewise for the predicate. Steinthal, for example, observed (1860: 101–102):

One should not be misled by the similarity of the terms. Both logic and grammar speak of subject and predicate, but only rarely do the logician and the grammarian speak of the same word as either the subject or the predicate.... Consider the sentence *Coffee grows in Africa*. There can be no doubt where the grammarian will locate subject and predicate. But the logician? I do not think the logician could say anything but that ‘Africa’ contains the concept that should be connected with ‘coffee grows’. Logically one should say, therefore, ‘the growth of coffee is in Africa’.

Observations like this gave rise to a long debate, which lasted more than 80 years. At the end, it was decided to keep the terms subject and predicate for the syntactic analysis and speak of topic and comment for the semantic analysis in the Aristotelian sense (see Seuren, 1998: 120–133 for a detailed discussion).

Syntax, in the modern sense, is largely absent from Aristotle’s writings. He does, however, distinguish between different sentence types (*On Interpretation* 17a1–12):

Every sentence is meaningful, not in virtue of some natural force but by convention. But not all sentences are assertions, only those in which there is question of truth or falsity. In some sentences that is not so. Wishes, for example, are sentences but they are not true or false. We will leave all other sentence types out of consideration, as they are more properly studied in rhetoric or poetics. But assertions are the topic of the present study [i.e., logic]. The primary assertive sentence type is the simple affirmation, the secondary is the simple negation. All other, complex, assertions are made one by

conjunction. Every assertion must contain a verb or a conjugated form of a verb. For a phrase like “man” is not yet an assertion, as long as no verb in the present, past, or future tense is added.

Some word classes are already there. Thus, at the outset of *On Interpretation*, he defines *ónoma* (noun) as “a stretch of sound, meaningful by convention, without any reference to time and not containing any internal element that is meaningful in itself” (*On Interpretation* 16a19–21). *Rhêma* (verb) is defined as “that which, in addition to its proper meaning, carries with it the notion of time, without containing any internal element that is meaningful in itself; it always is a sign of something said of something else” (*On Interpretation* 16b6–8). In his *Rhetoric*, at 1406a19, Aristotle uses the term *epíteton* for adjective. All other terms for word classes are of a later date, with many of them having been created by the Alexandrian linguists.

The term *ptôsis* is found relatively frequently, in the sense of nominal or verbal morphological modification, as in *Categories* 1a13–15: “Things are said to be named ‘derivatively’ when they derive their name from some other word that differs in morphological form (*ptôsei*), such as the grammarian from the word *grammar* or the courageous from the word *courage*.” The literal meaning of *ptôsis* is ‘fall’ (Latin: *casus*). Its use in the sense of morphological modification is based on the metaphor that the word ‘as such’ stands upright (in the ‘upright case’ or *orthê ptôsis*; Latin: *casus rectus*). Its other falls are represented by forms that are modified morphologically according to some paradigm. The Alexandrians began to reserve the term *ptôsis* for the nominal cases of *nominative* (the form of your own name), *genitive* (the form of your father’s name), *dative* (the name of the person you give something to), *accusative* (the name of the person you take to court), and *vocative* (the name of the person you call). These terms smell of the classroom, not of philosophy.

Bibliography

- Allan K (2004). ‘Aristotle’s footprints in the linguist’s garden.’ *Language Sciences* 26(4), 317–342.
- Ogden C K & Richards I A (1923). *The meaning of meaning. A study of the influence of language upon thought and of the science of symbolism*. London: Routledge & Kegan Paul.
- Seuren P A M (1998). *Western linguistics: An historical introduction*. Oxford: Blackwell.
- Steinthal H (1860). *Charakteristik der hauptsächlichsten Typen des Sprachbaues* (Neubearbeitung von Dr. Franz Misteli). Berlin: Dümmler.

Assertion

M S Green, University of Virginia, Charlottesville, VA, USA

© 2006 Elsevier Ltd. All rights reserved.

Plato in such works as *Cratylus* takes a sentence to be a series of names. This view is untenable because it is not the case that in uttering a series of names one thereby says something either true or false. “Ephesus, 483, Neptune” says nothing either true or false. By contrast, an indicative sentence, such as “Ephesus was a large city,” is true, and it seems to include elements other than names. Aristotle seems to be making just this point when in *De Interpretatione* he remarks

Falsity and truth have to do with combination and separation. Thus, names and verbs by themselves—for instance “man” or “white” when nothing further is added—are like the thoughts that are without combination and separation. So far they are neither true nor false (1963: 16^a9)

Aristotle, however, does not distinguish in his use of “statement” among (1) the sentence used to express a proposition, (2) the proposition thereby expressed, or (3) the use of that proposition in the performance of a speech act, such as an assertion. These are three distinct phenomena, but it took over two millennia before the point was formulated clearly by the philosopher-logician Gottlob Frege.

According to Frege, a proposition (such as the proposition that global warming is accelerating) is an abstract object, the existence of which does not depend upon any mind grasping it or any sentence expressing it (1984). By contrast, a sentence expresses a proposition if it is indicative and meaningful. Unlike propositions, sentences exist only as part of a conventional linguistic system and so in a clear sense depend upon human activities. Nevertheless, one can utter a sentence expressing a proposition without making an assertion. One might utter such a sentence in the course of testing a microphone, or in rehearsing one’s lines for a play, or for that matter in investigating a proposition to see what its consequences are without endorsing it. For instance, one might put forth the proposition that global warming will cause the melting of Greenland’s glaciers to see what would follow (the submersion of Florida, etc.) without claiming that global warming will in fact go so far as to melt Greenland’s glaciers.

There are three distinct items then: an indicative sentence, a proposition expressed by that sentence, and the use of that sentence and proposition

expressed to make an assertion. ‘Statement,’ ‘claim,’ ‘judgment,’ and even ‘proposition’ are often used interchangeably among these three notions, and in the history of philosophy no small amount of mischief has resulted from such ambiguous usage. (For a trenchant discussion of such ambiguity, see Geach, 1972; on the distinction between illocutionary force and semantic content, see Green, 2000.)

Isolating assertions from propositions and indicative sentences is only the beginning of wisdom about the former. An assertion is a speech act. Just as one can make a promise by saying, “I promise to do so and so,” so too one can assert P with such words as “I hereby assert that P.” (Austin, 1962 placed assertion on his list of ‘expositives’; Vanderveken, 1990 locates assertion on his list of ‘assertives.’) Further, just as a promise can only be made if certain conditions have been met (you cannot promise to do what is obviously outside your control), so too an assertion can only be made under certain conditions. For instance, it is doubtful that you can assert what is obvious to both you and your audience to be false. An attempt to do so will most likely be taken facetiously. Further, Strawson (1950) held that a sentence can make presuppositions that if not met prevent that sentence from being usable for assertions. If Iceland has no monarchy, then according to Strawson my utterance of, “The present queen of Iceland is beautiful,” will fail to make an assertion in spite of the fact that I have in good faith uttered a meaningful indicative sentence.

To assert a proposition is at the very least to put it forth as true. Searle (1969) tried to capture this dimension by construing an assertion of proposition P as an “undertaking to the effect that P.” That may be a necessary, but is not a sufficient condition: My betting a large sum on P’s coming out true is an undertaking to the effect that P is true, but it is no assertion thereof. (For instance, if I bet on P without believing P true, I am not a liar; if I assert P under those conditions, I am.) Searle, however, elsewhere wrote that an assertion has word-to-world direction of fit. According to this conception, inherent in the practice of asserting is the norm that the speaker’s words are supposed to track how things are. This feature distinguishes assertion from, say, commands, the aim of which is not to track the world but rather to get the world (in particular, one or more of its inhabitants) to conform to what the command enjoins. Commands are thus commonly said to have world-to-word direction of fit.

This notion of word-to-world direction of fit may be elaborated further by the way in which one sticks

one's neck out by making an assertion. One who asserts a proposition P is right or wrong on the issue of P , depending upon whether P is true: one is right on that issue if P is indeed the case and wrong if it is not. In thus exposing oneself to liability to error on the issue of P , one is doing more than if one had just uttered a sentence expressing P .

However, liability to error on the issue of a particular proposition still does not distinguish assertion from other speech acts involving propositions. One who guesses that P is also right or wrong on the issue of P , depending on whether P is the case. How may we distinguish assertion from other proposition-involving speech acts? Williamson (1996) contended that one who asserts P is thereby open to the challenge, "How do you know?" This much may not be said of other proposition-involving speech acts. For instance, it would be inappropriate to respond to one who conjectures, guesses, or supposes for the sake of argument that a black hole inhabits the center of the Milky Way, with the question, "How do you know?" Unlike these other speech acts, an assertion purports to convey knowledge.

One who makes an assertion incurs a risk through the aforementioned liability to error and, if Williamson is correct, is exposed to a conversational challenge. What if such a challenge is made? According to the view of Brandom (1983, 1994), in that case its issuer is obliged to respond with reasons that would justify the contested claim. Those reasons might invoke items of experience or the authority of others. The issuer of the assertion might even show that the speaker raising the challenge is committed to that proposition by her own convictions. If, however, the assertor is unable to convince the interlocutor of the assertion, others will become unable to defer to his or her authority if their own assertion of that same proposition is challenged.

In light of incurring a risk of error and exposing oneself to a conversational challenge, it might seem that asserting is more trouble than it is worth. Yet, asserting is the bread and butter of conversational life, so it presumably has redeeming points. First, one whose assertions turn out to be reliably correct, or at least widely accepted, garners credibility. That in turn is a source of social authority: We look to reliable assertors to get things right. Second, it might be held that knowledge has intrinsic worth. Although making an assertion that is borne out is not a sufficient condition for knowledge (you might have gotten lucky or not have believed what you claimed), it is often associated with it. For this reason, getting things right by means of a correct assertion might be thought to be its own reward.

Third, an assertion that is not challenged, or is challenged but the assertor responds with an adequate defense of the claim, may be entered into conversational common ground. A conversation begins with a (possibly empty) set of propositions that interlocutors hold in common while being mutually aware that they hold this information in common. Among fans at a baseball game this common ground might include propositions about current weather conditions, the teams' score, and perhaps a few items in the current national news. Much more will typically be common ground among members of a tightly knit family, less among people waiting at a major bus terminal (Clark, 1996).

Let Σ_i be a commitment store for interlocutor i , containing all those propositions to which interlocutor i is committed. Where $1, \dots, n$ are interlocutors, we may define $\Sigma_{1 \cap \dots \cap n}$ as $\Sigma_1 \cap \dots \cap \Sigma_n$. Even if $P \in \Sigma_{i \cap j}$, it does not follow that P is in the common ground of i and j , for neither may have any commitments with regard to the other's commitment to P . Rather, where s_1, \dots, s_n are a group of interlocutors,

P is common ground among s_1, \dots, s_n (written $P \in \Sigma_{1 \dots n}$) iff

- (a) for all $s_i \in \{s_1, \dots, s_n\}$, $P \in \Sigma_i$, and for all $s_i \in \{s_1, \dots, s_n\}$, (a) $\in \Sigma_i$.

$\Sigma_{1 \dots n}$ will in general be a proper subset of $\Sigma_{1 \cap \dots \cap n}$. When s_1, \dots, s_n are engaged in a conversation and s_i asserts that P , then so long as no other member of s_1, \dots, s_n demurs, $P \in \Sigma_{1 \dots n}$. This may bring about progress on a question at issue (Where is the bus station?; or Why is the child ill?) or may aid in the elaboration of a plan (of getting to the bus station or curing the child), and it enables speakers at a later time to presuppose P in their speech acts. For instance, once it is common ground that the child's illness is viral, we may presuppose this fact, as shown by the acceptability of a remark such as, "Since antibiotics will be useless on her, we need to ease her discomfort some other way." The use of the phrase, "Since antibiotics will be useless on her," would be inappropriate if it were not already common ground that the child's illness is viral and that antibiotics do not treat viral infections (see Green, 2000 for a fuller discussion).

In addition to enhancing inquiry and planning and to laying the foundation for later presuppositions, an assertion may generate other pragmatic effects. For instance, one who asserts P may also suggest, insinuate, or implicate a distinct proposition through the meaning of what one says: If one asserts that Mary was poor but honest, the use of "but" suggests that there is some tension between poverty and honesty.

According to Grice's theory of implicature (1989), in that case one does not, however, assert that there is such a tension. Rather, what one asserts is true only in case Mary is poor and Mary is honest. Again, if someone asserts that Mary lost a contact lens, he or she will normally be taken to suggest that Mary lost one of her own contact lenses, rather than someone else's contact lens. This implicature, due not to the conventional meaning of any particular locution but rather to norms governing conversational practice, is also no part of what he or she asserts. When a speaker asserts P, he or she often means (specifically, speaker means) a good deal more without thereby asserting those other contents.

Assertions are not just beholden to the norm of accuracy about the world; they are also held to the norm of sincerity. One who asserts what one does not believe is a liar. By contrast, it is no lie conversationally to implicate what one does not believe. If in response to your question, "Where is Mary?" I reply that she is somewhere in Spain, I may implicate that I am in no position to be more specific. If in fact I do know that she is in Pamplona, I have been evasive, misleading, and perhaps even mendacious but no liar (Adler, 1997). Similarly an assertion of P, while representing oneself as believing that P, is not also an assertion that one believes that P. Were that so, one would literally contradict oneself in saying, "P but I don't believe that P." As G. E. Moore (1993) observed, however, although this utterance does seem absurd in some way, it is not a self-contradiction. What it says could well be true. Although an assertion of P is not an assertion that one believes that P, that assertion does express, manifest, or display one's belief that P. Current research is actively engaged with the relation between assertion and the states of mind that it manifests (Williams, 1996; Davis, 2003; Green and Williams, 2006).

See also: Lying, Honesty, and Promising; Mood, Clause Types, and Illocutionary Force; Normativity; Pragmatic Determinants of What Is Said; Propositions; Truth: Primary Bearers.

Bibliography

- Adler J (1997). 'Lying, deceiving, or falsely implicating.' *Journal of Philosophy* 94, 435–452.
- Aristotle (1963). *De Interpretatione*. Acrrill J L (trans.). Oxford: Oxford University Press.
- Austin J L (1962). 'How to do things with words.' In Urmson J O & Sbisa M (eds.). Cambridge, MA: Harvard University Press.
- Brandom R (1983). 'Asserting.' *Noûs* 17, 637–650.
- Brandom R (1994). *Making it explicit*. Cambridge, MA: Harvard University Press.
- Davis W (2003). *Meaning, expression and thought*. Cambridge: Cambridge University Press.
- Carston R (2003). *Thoughts and utterances*. Oxford: Blackwell.
- Clark H (1996). *Using language*. Cambridge: Cambridge University Press.
- Frege G (1984). 'The thought: a logical inquiry.' In McGuinness B (ed.) *Collected papers on mathematics logic and philosophy*. Black M et al. (trans.). Oxford: Blackwell.
- Geach P (1972). 'Assertion.' In Geach P (ed.) *Logic matters*. Oxford: Blackwell. 254–269.
- Green M (2000). 'Illocutionary force and semantic content.' *Linguistics and Philosophy* 23, 435–473.
- Green M & Williams J (2006). *Moore's paradox: new essays on belief, rationality and the first person*. Oxford: Oxford University Press.
- Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- McDowell J (1998). 'Meaning, communication, and knowledge.' In *Meaning, knowledge and reality*. Cambridge, MA: Harvard University Press.
- Moore G E (1993). 'Moore's paradox.' In Baldwin T (ed.) *G. E. Moore: selected writings*. London: Routledge. 207–212.
- Plato (1998). *Cratylus*. Reeve C D C (trans.). Indianapolis: Hackett.
- Searle J (1969). *Speech acts*. Cambridge: Cambridge University Press.
- Strawson P (1950). 'On referring.' *Mind* 59, 320–344.
- Vanderveken D (1990). *Meaning and speech acts*. Cambridge: Cambridge University Press.
- Williamson T (1996). 'Knowing and asserting.' *The Philosophical Review* 105, 489–523.

B

Behaviorism: Varieties

P Poirier, Université du Québec à Montréal, Montréal, Québec, Canada

© 2006 Elsevier Ltd. All rights reserved.

The term ‘behaviorism’ refers to a family of doctrines that emphasize the importance of behavior over mind, or cognitive processing, in psychology, notably as its proper subject matter or its ultimate evidential basis.

Psychological Behaviorism

Early in the 20th century, James Watson wove together three 19th-century ideas – Darwin’s evolutionary theory emphasizing the physical as well as psychological continuity between animals and humans, Wundt’s experimental method in psychology, and James’s functionalist psychology – into both a method and theoretical overview for animal and human psychology. In 1913, he published what came to be known as the behaviorist manifesto, ‘*Psychology as the behaviorist views it*,’ which contained a number of distinct but related doctrines, three of which are of particular interest for the student of language: (1) the rejection of introspection as an experimental method, (2) the shunning of internal events, and (3) the emphasis on learning.

Rejection of Introspection as an Experimental Method

In the late 19th century, Wundt convinced psychologists that scientific work in their discipline must be experimental and quantitative, but like all 19th-century psychologists, he continued to work on psychology’s traditional object – consciousness – and to access it by the traditional route of introspection. Stirred by a growing methodological distrust in the value of introspective evidence, and emboldened by various successes in animal psychology in which introspective data were not available, Watson, trained as an animal experimenter, argued that consciousness and introspection were not a proper basis upon which to establish an experimental scientific psychology.

According to Watson, experimentation in psychology begins and ends with observable behavior, a methodological prescription that is sometimes called ‘methodological behaviorism.’ To this day, virtually all working psychologists, even those who see themselves as strong opponents of behaviorism, adhere to Watson’s methodological prescription. For this reason, it is better to reserve the term behaviorism for the other behavioristic doctrines present in Watson’s manifesto.

Shunning of Internal Events

If all internal events are characterized by the fact that they are conscious and accessible only through introspection, then Watson’s rejection of conscious experience as the object of psychology amounts to a rejection of all internal events from the scope of psychology. Indeed, we find in Watson’s writing a move from his initial 1913 doubts about the epistemic value of internal events (*qua* conscious events) to the full-fledged elimination of everything internal found in his 1924 monograph *Behaviorism*. However, the elimination of everything internal, a position we could call eliminative methodological behaviorism, which is one of the two main tenets most strongly identified with behaviorism today, is only one way to meet Watson’s methodological prescription. The other is to view internal events as theoretical constructs, the value of which flows from behavioral observations. According to this position, internal events can be posited as theoretical construct provided they are ‘operationally defined’ into behaviors. Few psychologists, even among those we would undeniably classify as behaviorists today, followed Watson’s shunning of internal events. Edward Tolman and Clark Hull, for instance, were much more ready to posit internal events (drives, internal maps, etc.) as intervening variables mediating the relation between stimuli and responses. Even Skinner found a heuristic place for internal events in his *Experimental Analysis of Behavior* (EAB), but reduced their epistemic value within psychology.

We can call this position, held by most behaviorists, ‘operationalist methodological behaviorism.’ In their

day-to-day practice, most psychologists today are operationalist methodological behaviorists. To see what distinguishes those operationalist methodological behaviorists we call behaviorists from those called, say, cognitive psychologists, one has to look at the third element found in Watson's behaviorist manifesto, the emphasis on learning.

Emphasis on Learning

Whereas the first two doctrines found in Watson's manifesto are methodological, the third is theoretical. The main construct on which Watson established his psychology is the reflex; indeed, he opened his 1924 monograph with an analysis of the yawning reflex. However, Watson recognized that basic reflexes, which are fixed and innate stimulus-response (S-R) dispositions, have insufficient explanatory power to account for the behavior of many animals, and certainly those that interested Watson: birds and mammals, including humans.

To achieve the necessary explanatory power, Watson turned to the work of Russian physiologist Ivan Pavlov (1927), who had recently won the Nobel Prize in medicine for his work on a type of learning called classical conditioning. Pavlov showed that, when the temporal contingencies are right, stimuli that are neutral to the organism can be paired with innate (or unconditioned) S-R dispositions so as to acquire the ability to trigger the reflexes' response. Once the new association is established, there is a new, conditioned S-R disposition in the organism's behavioral repertoire. Since so many of human behavioral dispositions are not innate, Watson hoped that classical conditioning would offer a means to account for what is typically human in human psychology. Note that Watson's emphasis on a learning mechanism does not make him an empiricist: indeed, for his own brand of behaviorism to get off the ground, he needs to posit a large number of innate unconditioned reflexes. With the development of the concept of operant conditioning, and to the extent he relegated classical conditioning to a secondary role, Skinner was actually much more of an empiricist than Watson ever could be.

Skinner developed both a school of psychological research – the Experimental Analysis of Behavior – and a philosophy to underwrite it, called radical behaviorism (Skinner, 1938, 1974). According to EAB, every behavior can be broken down into three parts: the discriminative stimulus (the type of situation the organism is in), the operant response (a response that can be modified as a function of its consequence), and another stimulus: the reinforcer

or punisher – an event that increases (reinforcement) or decreases (punishment) the probability that the organism will produce the operant response in the presence of the discriminative stimulus. Operant conditioning, a general learning mechanism, is said to have occurred when a stimulus has increased or decreased the probability that an organism will produce a given operant response when the discriminative stimulus is present.

Unlike Watson's elimination of inner mental states, Skinner's EAB did not shy away from them, but viewed them as operant responses, which could be reinforced or punished by stimuli. Contrary to psychological tradition, however, inner mental states were not seen as causes of behavior, but rather as behaviors (operant responses) in their own right, to be controlled by external stimuli like any other operant response. And unlike classical conditioning – the learning mechanism favored by Watson – operant conditioning is a *creative process*, much the same way natural selection is, allowing for the selection of new responses. Armed with operant conditioning and classical conditioning to a lesser extent, Skinner believed he could account for any human behavior.

In his discourse on method, Descartes famously set a limit on mechanistic explanation in psychology. No mere mechanism could, he believed, account for the productivity and pragmatic appropriateness of language and reasoning. In his 1957 work, *Verbal Behavior*, Skinner sought to show that, by applying the concepts and methods of EAB, one could mechanistically account for both the productivity and pragmatic appropriateness of language. His account of pragmatic appropriateness took sentences as wholes and showed how the environment, especially the verbal community, provides reinforcers and punishers to shape verbal responses. His account of productivity focused on words and attempted to show how sentences are structured by reinforcers and punishers provided by the child's verbal community.

In a famous review of Skinner's book, Chomsky (1959) attacked Skinner's behaviorist account of language. Chomsky set up a two-horn dilemma on which Skinner could only impale himself. Either Skinner uses EAB's theoretical terms in a strict and controlled way – that is, as he defined them operationally – or he interprets them in a loose and metaphorical way. If he interprets its terms in a loose and metaphorical way, then his analysis of language amounts to nothing but a disguised form of mentalism. On the other hand, if Skinner interprets EAB's theoretical terms as he defined them operationally, then, according to Chomsky, Skinner's behaviorism could not account for language.

Operant conditioning shapes verbal behavior by differentially reinforcing and punishing verbal output. One can easily see how this process could account for the pragmatic appropriateness of verbal responses in anyone who can produce them. (Chomsky did not emphasize this aspect of verbal behavior, perhaps because he himself did not have an account of pragmatic appropriateness.) However, according to Chomsky's famous "poverty of stimulus argument," operant conditioning cannot account for the acquisition of the capacity to produce grammatical verbal responses. The linguistic environment (the stimulus) is said to provide children with insufficient evidence – to be precise, evidence that was too *unsystematic* and too weak – to acquire the grammar of the language spoken around them: unsystematic because it contains no negative evidence (evidence that identifies ungrammatical verbal output as ungrammatical) and *weak* because positive evidence (evidence that identifies grammatical verbal output as grammatical) is of low quality. Moreover, because children hear only a small finite subset of all the grammatical sentences of their language, they cannot take the absence of a grammatical form as evidence of ungrammaticality. Yet, despite the poor nature of the stimulus they get from their verbal environment, children normally acquire their native language quite rapidly.

Over the years, every element of this argument has been argued extensively. Skinner himself never bothered to reply to Chomsky since it was obvious to him that Chomsky had not read his book. As noted, *Verbal Behavior* did in many ways offer an account of language's pragmatic appropriateness (one of Descartes' challenges to a mechanistic explanation of the mind), something that formal grammars cannot. However, Chomsky was right in claiming that Skinner's book offered no clue to the productivity of language (Descartes' other challenge), whereas his own did. One could say in retrospect that the score between Skinner and Chomsky is one-all. Nevertheless, by the 1960s, the tide had turned against behaviorism, and the time was ripe for cognitive science. Chomsky's attack on Skinner's *Verbal Behavior* was seen as one of the definitive moments in the birth of cognitive science.

Philosophical Behaviorism

Philosophical behaviorism is a semantic theory about the meaning of mental statements; that is, statements used to attribute mental states to oneself or another. Although philosophical behaviorism generally is often identified with the particular brand developed

by neopositivists, it is important to note that it has three main strands. They are, chronologically, the reductionist philosophical behaviorism of the Vienna Circle and two related but distinct responses to it: the nonreductionist philosophical behaviorism of ordinary language philosophers and the eliminativist philosophical behaviorism of Quine. All of these positions are behaviorist in that they posit a special relation between behavior or behavioral statements and the meaning of mental statements. What distinguishes the three is the type of relation they posit.

Reductive Philosophical Behaviorism

Developed by members of the Vienna Circle as part of their unity-of-science program, reductionist philosophical behaviorism (also called logical or analytical behaviorism) posits that there can be semantic identity between mental statements and (a finite subset of) behavioral statements. Two specific neopositivist theses are at the heart of this approach: (1) positivism – the rejection, as meaningless, of metaphysics and pseudosciences from the unified system of the human knowledge (science) and (2) atomistic verificationism, the view that theoretical statements, taken individually, draw their meaning from their relation of semantic identity to special foundational statements, called 'physical control formulas,' the truth of which, individually, is taken to be directly verifiable by observation. Note that, although atomistic verificationists (and many contemporary commentators) use terms like 'synonymy' to characterize the type of semantic relation posited and 'translation' to characterize the operation carried on to establish the semantic relation, neither terms should be confused with its homonym in ordinary language. Rather, the relation of semantic identity posited by atomistic verificationists is akin to the relations of definition in mathematics or of compilation in computer science (both are relations between elements in formal languages).

Anyone who holds these two views, and believes in the value of psychology as a science will strive to show that psychological statements, as theoretical statements, are semantically identical to physical control formulas: physiological and behavioral statements. Hence, Hempel claimed that 'Paul has a toothache' was semantically identical to 'Paul weeps and makes gestures of such and such kinds'; "At the question 'What is the matter?' Paul utters the words 'I have a toothache'"; and so on (Hempel, 1949: 17).

Just as cognitive science was in part developed as a reaction to the limits of psychological behaviorism, in particular Skinner's EAB, much of the original philosophical background underlying cognitive science

was developed through a thorough reflection on the limits of philosophical behaviorism. Indeed, three broad philosophical views on the mind emerged from a reflection on the limits of reductionist philosophical behaviorism: nonreductionist philosophical behaviorism, eliminativist behaviorism, and functionalism. This article addresses the first two (*see Functional Theories of Language* for the third).

Nonreductive Philosophical Behaviorism

Whereas neopositivists aimed to show that mentalist psychology could be a part of the unified system of science, ordinary language philosophers wanted to understand the day-to-day use of language, including mentalistic language (language that contains mental statements and thus can be used to make mental attributions). First, ordinary language philosophers observed that mental attributions are, first and foremost, a mundane practice and that mentalist language may be ill suited to a scientific endeavor such as psychology. Simply assuming that mental statements were well suited for scientific work because of their value in ordinary human practices was deemed to be a category mistake (Ryle, 1949), or as Wittgenstein put it, to misconstrue the logical grammar of language (Wittgenstein, 1953).

The main error thus made was to assume that mental statements ascribe ‘states,’ presumably mental ones, but the criticism extended to other types of states (neurological, functional) of individuals. Instead, mental statements attribute complex behavioral dispositions or behavioral patterns to individuals. Such patterns are taken to be so complex that mental statements cannot in practice be reduced to them. Thus, according to nonreductive philosophical behaviorists, when we attribute a mental state to someone, we are not ascribing a given internal state to that person, but are using shorthand to describe what the person might do in various circumstances.

Ordinary language philosophers based their opposition to the brand of philosophical behaviorism developed by neopositivists on two related problems they saw with reductionist philosophical behaviorism. The first was that reductionist philosophical behaviorists conflated two distinct notions under the heading ‘behavior’: sometimes, the term referred to movements (of the body), and sometimes it referred to actions (of the person who has that body). Although movement statements can be properly viewed as physical control formulas, action statements cannot. The reason why action statements cannot be viewed as physical control formulas pointed to the second problem that philosophical behaviorists saw with reductionist philosophical behaviorism: the

intentional circle (Geach, 1957). The intentional circle refers to the fact that a given mental statement is semantically identical to a given action statement (or set thereof) only on the condition that the person the mental state is attributed to also holds other mental states, and that, to attribute those, one would use mental statements themselves semantically identical to actions only on the condition that the person holds other mental states; and so on, indefinitely. Psychological theoretical statements are thus shown to be irreducible to physical control formulas. The intentional circle argument is a particular instance of a more general argument that applies to any attempt to use semantic identity to justify the reduction of theoretical statements individually to physical control formulas. Let us turn to that more general argument.

Eliminative Philosophical Behaviorism

Atomistic verificationism is at the heart of the neopositivist conception of science, including its reductive philosophical behaviorism. The intentional circle argument brings into question the idea that psychological theoretical statements can be translated solely into physical control formulas – that is, without referring back to other psychological theoretical statements. It is as if all statements, theoretical and observational, are a connection in a kind of network. Drawing on Pierre Duhem’s confirmation holism, Quine broadened the scope of the intentional circle argument to cover all of the sciences: confirmational circularity between theoretical and observational statements becomes the norm rather than psychology’s exception. Holding on to verificationism, which ties meaning to observational confirmation, Quine rejected atomistic verificationism in favor of holistic verificationism. Behaviorism in psychology thus stops being a peculiarity in the sciences, owing to the quirky nature of its theoretical statements, but equal in that regard to theories in other sciences.

Having thus broadened the ordinary language philosopher’s argument against reductive philosophical behaviorism, Quine could have opted for their nonreductive behaviorism. However, he supplied his argument against reductive behaviorism with an argument against all intensional language in science; that is, all language in which extensional inference principles, such as existential generalization and the intersubstitutability of identicals, fail. Nevertheless, Quine’s position was quite similar to that of the nonreductive behaviorists. They both rejected reductive behaviorism and mentalist scientific psychology: Quine, because of his rejection of intensional language generally, and ordinary language philosophers

because they viewed with suspicion any extension of ordinary linguistic practices to science.

The Return of Psychological Behaviorism?

Most accounts of psychological behaviorism end with Chomsky's successful attack on Skinnerian behaviorism, and most accounts of philosophical behaviorism end with Quine's broad attack on neo-positivist reductive physicalism, forgetting that he replaced reductive behaviorism with a more radical eliminative behaviorism. However, since most linguists and psychologists did not follow in Quine's eliminative footsteps, opting instead for the then-growing paradigm of cognitive science, which promised a science of intentionality, a modern student of psychology and linguistics may thus be led to believe that behaviorism is long dead and buried.

However, various elements of behaviorist thinking are making a comeback. Eric Kandel won the Nobel Prize in Medicine in 2000 for his discovery of the molecular basis of three related forms of learning: sensitization, habituation, and classical conditioning (see Kandel *et al.*, 2000). In addition, many neural modelers are now replacing back-propagation in their connectionist models of cognitive capacities with reinforcement learning algorithms, inspired by Skinnerian learning mechanisms. Those algorithms have certain computational advantages over back-propagation that make them more suitable for learning in complex systems, and they are thought to be much more biologically plausible than back-propagation. Moreover, the physiological mechanisms of reinforcement learning are currently being worked out and are thought to involve dopamine and dopaminergic systems as reward systems. Finally, a new trend in robotics, suggestively called Behavior-Based Robotics, shuns central representations and computational systems in favor of simple behaviors strategically connected in a subsumption architecture (Brooks, 1999).

The two questions to ask, a half-century after the initial rejection of behaviorism by cognitive science, is not whether behaviorism is false (it is) but (1) to what extent does our intelligence rely on reflex-like input-output or a modular mechanism, and (2) to what extent is our cognitive life shaped by the type of learning mechanisms studied by behaviorists?

See also: Functionalist Theories of Language.

Bibliography

- Brooks R (1999). *Cambrian intelligence*. Cambridge, MA: MIT Press.
- Chomsky N (1959). 'A review of B. F. Skinner's Verbal Behavior.' *Language* 35(1), 26–58.
- Geach P (1957). *Mental acts: their content and their objects*. London: Routledge & Kegan Paul.
- Hempel C (1949). 'The Logical Analysis of Psychology.' In Feigl H & Sellars W (eds.) *Readings in philosophical analysis*. New York: Appleton-Century-Crofts. 373–384.
- Kandel E R, Schwartz J H & Jessell T M (2000). *Principles of neural science*. New York: McGraw-Hill.
- Pavlov I P (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford: Oxford University Press.
- Quine W (1953). *From a logical point of view*. Cambridge: Harvard University Press.
- Ryle G (1949). *The concept of mind*. London: Hutchinson.
- Skinner B F (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
- Skinner B F (1953). *Science and human behavior*. New York: Macmillan.
- Skinner B F (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Skinner B F (1974). *About behaviorism*. New York: Knopf.
- Watson J B (1913). 'Psychology as the behaviorist views it.' *Psychological Review* 20, 158–177.
- Watson J B (1924). *Behaviorism*. New York: Norton.
- Wittgenstein L (1953). *Philosophical investigations*. Anscombe G E M (trans.). Oxford: Blackwell.

Further Reading

- Carnap R (1932). 'Psychology in Physical Language.' *Erkenntnis* 3, 107–142.
- Hull C (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.
- Lewis D K (1980). 'Mad pain and martian pain.' In Block N (ed.) *Readings in philosophy of psychology*. Cambridge, Mass: Harvard University Press.
- Putnam H (1967). 'The nature of mental states.' In Putnam H (ed.) (1975) *Mind, language, and reality*. Cambridge: Cambridge University Press.
- Quine W (1960). *Word and object*. Cambridge, MA: MIT Press.
- Skinner B F (1948). *Walden two*. New York: Macmillan.
- Skinner B F (1971). *Beyond freedom and dignity*. New York: Knopf.
- Thorndike E. *Animal Intelligence*. Online: <http://psychclassics.yorku.ca/Thorndike/Animal/chap5.htm>. First published 1911.
- Titchener E B (1914). 'On "Psychology as the Behaviorist Views it".' *Proceedings of the American Philosophical Society* 53, 1–17.
- Tolman E (1948). 'Cognitive maps in rats and men.' *Psychological Review* 55, 189–208.

Boole and Algebraic Semantics

E L Keenan, University of California, Los Angeles, CA, USA

A Szabolcsi, New York University, New York, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

In 1854 George Boole, a largely self-educated British mathematician, published a remarkable book, *The laws of thought*, in which he presented an algebraic formulation of “those operations of the mind by which reasoning is performed” (Bell, 1965: 1). Since then, boolean algebra has become a rich subbranch of mathematics (Koppelberg, 1989), with extensive applications in computer science and, to a lesser extent, linguistics (Keenan and Faltz, 1985). Here we illustrate the core boolean notions currently used in the study of natural language semantics. Most such applications postdate Boole’s work by more than a century, though Boole (1952: 59) anticipated some of the linguistic observations, pointing out, for example, that *Animals are either rational or irrational* does not mean the same as *Either animals are rational or animals are irrational*; similarly, *Men are, if wise, then temperate* does not mean *If all men are wise then all men are temperate*. Generative grammarians rediscovered such truths in the latter third of the 20th century.

We begin with the basic notion of a *partially ordered set* (poset) and characterize richer structures with linguistic applications as posets satisfying additional conditions (Szabolcsi, 1997; Landman, 1991).

A poset consists of a domain D of objects on which is defined a binary relation R , called a *partial order* relation, which is *reflexive* (for all x in D , xRx), *transitive* (xRy and yRz implies xRz), and *antisymmetric* (xRy and yRx implies $x = y$). For example, the ordinary arithmetical \leq relation is a partial order: $n \leq n$, any natural number n ; if $n \leq m$ and $m \leq p$, then $n \leq p$; and if $n \leq m$ and $m \leq n$, then $n = m$. Similarly, the subset relation \subseteq is reflexive: any set A is a subset of itself. And if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$, so \subseteq is transitive. And finally, if $A \subseteq B$ and $B \subseteq A$, then $A = B$, that is, A and B are the same set, since they have the same members. So partial order relations are quite familiar from elementary mathematics.

A case of interest to us is the arithmetical \leq restricted to $\{0, 1\}$. Here $0 \leq 1$, $0 \leq 0$ and $1 \leq 1$, but 1 is not ≤ 0 . Representing the truth value ‘False’ as 0 and ‘True’ as 1, we can say that a conditional sentence ‘if P then Q ’ is True if and only if $TV(P) \leq TV(Q)$, where $TV(P)$ is the truth value of P , etc. Thus we think of sentences of the True/False sort as denoting in a set $\{0, 1\}$ on

which is defined a partial order, \leq . The denotations of expressions in other categories defined in terms of $\{0, 1\}$ inherit this order. For example, one-place predicates ($P1$ s), such as *is even* or *lives in Brooklyn*, can be presented as properties of the elements of the set E of objects under discussion. Such a property p looks at each entity x in E and says ‘True’ or ‘False’ depending on whether x has p or not. So we represent properties p, q as functions from E into $\{0, 1\}$, and we define $p \leq q$ if and only if (iff) for all x in E , $p(x) \leq q(x)$, which just means if p is True of x , then so is q . The \leq relation just defined on functions (from E into $\{0, 1\}$) is provably a partial order.

Other expressions similarly find their denotations in a set with a natural partial order (often denoted with a symbol like ‘ \leq ’). A crucial example for linguists concerns the denotations of count NPs (Noun Phrases), such as *some poets*, *most poets*, etc., as they occur in sentences (S s) like *Some poets daydream*. We interpret this S as True iff there is an entity x that both the ‘poet’ property p and the ‘daydreams’ property d map to 1. Similarly, *No poets daydream* is True iff there is no such x . And *Most poets daydream* is True iff the set of x such that $p(x)$ and $d(x) = 1$ outnumbers the set such that $p(x) = 1$ and $d(x) = 0$. That is, the set of poets that daydream is larger than the set that don’t. And for E, G possible NP denotations (called *generalized quantifiers*), we define $F \leq G$ iff for all properties p , $F(p) \leq G(p)$. This relation is again a partial order.

As NP denotations map one poset (properties) to another (truth values), it makes sense to ask whether a given function F *preserves* the order (if $p \leq q$, then $F(p) \leq F(q)$), *reverses* it (if $p \leq q$, then $F(q) \leq F(p)$), or does neither. *Some/all/most poets* preserve the order, since, for example, *is laughing loudly* \leq *is laughing* and *Some poet is laughing loudly* \leq *Some poet is laughing*, which just means, recall, that if the first sentence is True, then the second is. In contrast, *no poet* reverses the order, since, in the same conditions, *No poet is laughing* implies *No poet is laughing loudly*. The reader can verify that *fewer than five poets*, *neither poet*, *at most six poets*, and *neither John nor Bill* are all order reversing. And here is an unexpected linguistic correlation: reversing order correlates well with those subject NPs that license negative-polarity items, such as *ever*:

- (1a) No student here has ever been to Pinsk.
- (1b) *Some student here has ever been to Pinsk.

Observe that as a second linguistic application, modifying adjectives combine with property-denoting expressions (nouns) to form property-denoting

expressions and can be represented semantically by functions f from properties to properties. For example, *tall* combines with *student* to form *tall student*, and semantically it maps the property of being a student to that of being a tall student. And overwhelmingly when f is an adjective function and p a property, $f(p) \leq p$. All tall students are students, etc.

In fact, the denotation sets for the expressions we have discussed possess a structure much richer than a mere partial order: they are (boolean) lattices. A *lattice* is a poset in which for all elements x, y of the domain, the set $\{x, y\}$ has a *least upper bound* (lub) noted $(x \vee y)$ and read as ‘ x join y ,’ and a *greatest lower bound* (glb), noted $(x \wedge y)$ and read as ‘ x meet y .’ An *upper bound* (ub) for a subset K of a poset is an element z that every element of K is \leq to. An ub z for K is a lub for K iff $z \leq$ every ub for K . Dually a *lower bound* (lb) for K is an element $w \leq$ every element of K ; such a w is a glb for K iff every lb for K is $\leq w$. For example, in the truth value lattice $\{0,1\}$, lubs are given by the standard truth table for disjunction: $1 \vee 1 = 1$, $1 \vee 0 = 1$, $0 \vee 1 = 1$, and $0 \vee 0 = 0$. That is, a disjunction of two false Ss is False, but True otherwise. Similarly, glbs are given by the truth table for conjunction: a conjunction of Ss is True iff each conjunct is, and False otherwise. So here the denotation of *or* is given by \vee , and that for *and* by \wedge . And this is quite generally the case. In our lattices of functions, for example, $f \vee g$, the lub of $\{f, g\}$, is that function mapping each argument x to $f(x) \vee g(x)$. Similarly, $f \wedge g$ maps each x to $f(x) \wedge g(x)$. So, for example, in the lattice of properties, the glb of $\{\text{POET}, \text{DOCTOR}\}$ is that property which an entity x has iff $\text{POET}(x) = 1$ and $\text{DOCTOR}(x) = 1$, that is, x is both a poet and a doctor. So, in general, we see that the lattice structure provides denotations for the operations of conjunction and a disjunction, regardless of the category of expression we are combining. We might emphasize that the kinds of objects denoted by Ss, P1s, Adjectives, NPs, etc., are quite different, but in each category conjunctions and disjunctions are generally interpreted by glbs and lubs of the conjuncts and disjuncts. So Boole’s original intuition that these operations represent properties of mind – how we look at things – rather than properties specific to any one of these categories, is supported.

And we are not done: boolean lattices present an additional operation, *complement*, which provides a denotation for negation. Note that negation does combine with expressions in a variety of categories: with Adjectives in *a bright but not very diligent student*, with P1s in *Most of the students drink but don’t smoke*, etc. Formally, a lattice is said to be *bounded* if its domain has a glb (noted 0) and a lub (noted 1). Such a lattice is *complemented* if for every x there is a y such

that $x \wedge y = 0$ and $x \vee y = 1$. If for each x there is exactly one such y , it is noted $\neg x$ and called the *complement* of x . In $\{0, 1\}$, for example, $\neg 0 = 1$ and $\neg 1 = 0$. In our function lattices, $\neg f$ is that function mapping each x to $\neg(f(x))$. In distributive lattices (ones satisfying $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ and $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$), each x has a unique complement. A lattice is called *boolean* if it is a complemented distributive lattice. And, again, a linguistic generalization: the negation of an expression d in general denotes the complement of the denotation of d .

Given uniqueness of complements, \neg is a function from the lattice to itself, one that reverses the order: if $x \leq y$, then $\neg y \leq \neg x$. We expect, correctly then, that negation licenses negative-polarity items in the predicate, and it does: *He hasn’t ever been to Pinsk* is natural, **He has ever been to Pinsk* is not. Reversing the order on denotations, then, is what ordinary negation has in common with NPs such as *no poet*, *neither John nor Bill*, etc., which as we saw earlier also license negative-polarity items.

The boolean lattices we have so far invoked have further common properties. They are, for example, *complete*, meaning that each subset, not just ones of the form $\{x, y\}$, has a glb and a lub. They are also atomic (Keenan and Faltz, 1985: 56). In addition, different categories have some distinctive properties – which, with one exception, space limitations prevent us from reviewing (see also Keenan, 1983). The exception is the lattice of count NP denotations, needed for expressions such as *most poets* and *five of John’s students*. This lattice has the property of having a set of complete, independent (free) generators, called *individuals* (denotable by definite singular NPs, such as *John*, *Mary*, *this poet*). This means that any function from properties to truth values is in fact a boolean function (meet, join, complement) of individuals (Keenan and Faltz, 1985: 92). And this implies that the truth value of an S of the form $[[\text{Det N}] + \text{P1}]$, for P1 noncollective, is booleanly computable if we know which individuals have the N and the P1 properties. The truth of Ss like *Most of the students laughed*, *No students laughed*, etc., is determined once that information is given. This semantic reduction to individuals is a major simplification, in that the number of individuals is the number of elements in E , whereas the number of possible NP denotations is that of the power set of the power set of E . So speaking of an E with just four elements, we find there are just four individuals but 65 536 NP denotations.

These freely generated algebras show up in another, unexpected syntactic way. Szabolcsi and Zwarts (1993) observed that negation determines a context that limits the class of questions (relative clauses, etc.) we can grammatically form. Thus, the questions

in (2) are natural, but those in (3), in which the predicates are negated, are not:

- (2) How tall is John? How much did the car cost?
 (3) *How tall isn't John? *How much didn't the car cost?

It is tempting to say simply that we cannot question out of negative contexts, but that is not correct. Both questions in (4) are acceptable:

- (4) How many of the books on the list did/didn't you read?

A more accurate statement is that negation blocks questioning from domains that lack individuals (free generators), such as amounts and degrees. So, as with the distribution of negative-polarity items, we find an unexpected grammatical sensitivity to boolean structure.

Much ongoing work in algebraic semantics focuses on NPs (and their predicates) that are not boolean compounds of individuals. The predicates in the Ss in (5) force us to interpret their subjects as *groups*.

- (5a) John and Mary respect each other/are a nice couple.
 (5b) Russell and Whitehead wrote *Principia mathematica* together.
 (5c) The students gathered in the courtyard/surrounded the building.
 (5d) Six teaching assistants graded 120 papers between them.

Respecting each other (*being a nice couple*, etc.) holds of a group of individuals if certain conditions among them obtain. But it does not make sense to say **John respects each other* (**He is a nice couple*, etc.), so we must interpret *and* somewhat differently from the glb operator discussed earlier. We note that the other boolean connectives – such as *either ... or ...* and *neither ... nor ...* – do not admit of a reinterpretation in the way that *and* does (Winter, 2001). **Either John or Mary respect each other* is nonsense: the disjunctive subject still forces a lub interpretation in which *respect each other* would hold of at least one of the disjuncts.

First attempts to provide denotations for the subject NPs in (5) involve enriching the understood domain E of entities with a partial order relation called *part-of*, to capture the sense in which the individual John is part of the denotation of *John and Mary* in (5a) or some individual student is part of the group of students in (5c), etc. The group itself is a new type of object, one that is the lub of its parts. And new types of predicates, such as those in (5), can select these new objects as arguments. Thus, the domain of a model is no longer a mere set E but is a *join semi-lattice*, a set equipped with a *part-of* partial order in which each

nonempty subset has a lub (see Link, 1983, 1998; Landman, 1991).

Yet other new types of arguments are *mass terms* (6a) and *event nominals* (6b).

- (6a) Water and alcohol don't mix.
 (6b) 4000 ships passed through the lock last year.
 (Krifka, 1991)

Mass term denotations have a natural *part-of* relation: if I pour a cup of coffee from a full pot, the coffee that remains, as well as that in my cup, is part of the original coffee. So mass term denotations are in some way *ontologically uniform*, with the result that definitional properties of a whole also apply to their parts – the coffee I poured and the coffee that remains are both coffee. This contrasts with predicates in (5), where *respect each other*, *gather in the courtyard*, etc., do not make sense even when applied to the proper parts of their arguments. In general, mass terms are much less well understood than count terms (see Pelletier and Schubert, 1989; Link, 1998).

Last, observe that (6b) is ambiguous. It has a count reading, on which there are 4000 ships each of which passed through the lock (at least once) last year. But it also has an event reading, of interest here, on which it means that there were 4000 events of ships passing through the lock. If, for example, each ship in our fleet of 2000 did so twice, then there were 4000 passings but only 2000 ships that passed.

Now, the event in (6b) has the individual passing events as parts, so such complex events exhibit something of the ontological uniformity of mass terms. But there are limits. The subevents of a single passing (throwing lines to the tugboats, etc.) are not themselves passings. So events present a *part-of* partial order with limited uniformity, and at least some events can be represented as the lubs of their parts. But in distinction to pure mass terms, events are ontologically complex, requiring time and place coordinates, Agent and Patient participants, etc., resulting in a considerable enrichment of our naïve ontology (see Parsons, 1990; Schein, 1993; and Landman, 2000).

See also: Formal Semantics; Monotonicity and Generalized Quantifiers; Negation: Semantic Aspects; Plurality; Quantifiers: Semantics.

Bibliography

- Bell E (1937). *Men of mathematics*. New York, NY: Simon and Schuster.
 Boole G (1854). *The laws of thought*. Reprinted (1952) as vol. 2 in *George Boole's collected logical works*. La Salle, IL: Open Court.

-
- Carlson G (1977). 'A unified analysis of the English bare plural.' *Linguistics and Philosophy* 1, 413–456.
- Keenan E L (1983). 'Facing the truth: some advantages of direct interpretation.' *Linguistics and Philosophy* 6, 335–371.
- Keenan E L & Faltz L M (1985). *Boolean semantics for natural language*. Dordrecht: D. Reidel.
- Koppelberg S (1989). Monk J D & Bonnet R (eds.) *Handbook of boolean algebras, vol. 1*. North-Holland: Amsterdam.
- Krifka M (1991). 'Four thousand ships passed through the lock: object-induced measure functions on events.' *Linguistics and Philosophy* 13, 487–520.
- Krifka M (1992). 'Thematic relations as links between nominal reference and temporal constitution.' In Sag I A & Szabolcsi A (eds.) *Lexical matters*. Chicago: CSLI Publications, Chicago University Press. 29–53.
- Landman F (1991). *Structures for semantics*. Dordrecht: Kluwer.
- Landman F (2000). *Events and plurality*. Dordrecht: Kluwer.
- Link G (1983). 'A logical analysis of plurals and mass terms: a lattice-theoretic approach.' In Bäuerle R *et al.* (eds.) *Meaning, use and interpretation in language*. Berlin: de Gruyter. 302–323.
- Link G (1998). *Algebraic semantics in language and philosophy*. Stanford: CSLI.
- Parsons T (1990). *Events in the semantics of English: a study in subatomic semantics*. Cambridge, MA: MIT Press.
- Pelletier F J & Schubert L K (1989). 'Mass expressions.' In Gabbay D & Guenther F (eds.) *Handbook of philosophical logic, vol. IV*. Dordrecht: D. Reidel. 327–407.
- Schein B (1993). *Plurals and events*. Cambridge, MA: MIT Press.
- Szabolcsi A (ed.) (1997). *Ways of scope taking*. Dordrecht: Kluwer.
- Szabolcsi A & Zwarts F (1993). 'Weak islands and an algebraic semantics for scope taking.' *Natural Language Semantics* 1, 235–284.
- Winter Y (2001). *Flexibility principles in boolean semantics*. Cambridge, MA: MIT Press.
-

This page intentionally left blank



Causal Theories of Reference and Meaning

A Sullivan, Memorial University of Newfoundland,
St. John's NL, Canada

© 2006 Elsevier Ltd. All rights reserved.

ismuch more important on the causal approach to reference, as compared with the causal approach to meaning.)

Reference, Meaning, and Causal Theories

The theory of reference and the theory of meaning are two closely related, fundamental strains within the study of mind and language. The aim of a theory of meaning is to explain what it is that competent speakers of a given language know, or are able to do, in virtue of which they are able to use the language to communicate. The aim of the theory of reference is to explain what it is in virtue of which words refer to what they do, how it is that utterances can hook onto and express information about particular things.

The exact relation between meaning and reference is a controversial matter (in large part because of the wide variety of theoretical approaches to meaning). According to some views, the meaning of an expression is precisely its referent, and so theories of meaning and of reference are just slightly different roads in to what is essentially the same task. Opponents of this notion point to co-referential expressions that differ in meaning (such as 'Portugal' and 'the country immediately west of Spain'), or to meaningful expressions that do not seem to refer to anything ('of', or 'for the sake of'), to show that meaning is distinct from reference. Or again, many theorists hold that proper names refer but cannot really be said to have a meaning, or that complete sentences have a determinate meaning but do not refer to anything.

In any case, the causal theory of reference (i.e., words refer to what they do by virtue of a certain sort of causal relation between word and referent) and the causal theory of meaning (i.e., words mean what they do by virtue of a certain sort of causal relation between word and meaning) are, historically and conceptually, distinct views. To help avoid confusion, I will distinguish the relevant approach to reference by calling it 'the causal-historical theory.' ('Historical' is an appropriate distinguishing mark because the history of how a word is transmitted from its original inception to the current speaker

The Causal-Historical Theory of Reference

The causal-historical theory of reference was developed in the 1960s and 1970s. It is explicitly developed only for proper names (cf. Donnellan, 1970; Kripke, 1972) and natural kind terms (cf. Kripke, 1972; Putnam, 1975). However, Kaplan (1977) raises some related points about indexical expressions, and there have been attempts to fashion a fully general approach to reference along these lines (for discussion, see Stalnaker, 1997; Devitt and Sterenly, 1999). The theory has replaced the descriptivist approach to reference, different versions of which were defended by Frege and Russell, as the orthodox approach to reference. (see **Proper Names: Philosophical Aspects** for discussion.)

According to the causal-historical theorists, descriptivists are wrong to demand that, in order to significantly use a term, speakers need to have a uniquely identifying description of its referent. Rather, once a convention is in place, linking a term to its referent, a deferential intention to comply with this practice – i.e., to use 'X' to refer to what others have used 'X' to refer to – is all that is required in order to use the term to refer. The view has it that certain expressions refer to certain things in virtue of a causal-historical relation between word and object, initially fixed during a dubbing or baptism and propagated from there to subsequent speakers, who implicitly defer to that initial dubbing in using the expression to refer.

The notion of a causal-historical chain as that which is criterial in determining reference is developed more or less independently by Donnellan and Kripke. Donnellan (1970: 277) concludes an argument against descriptivism with the claim that "... in some way the referent must be historically, or, we might say, causally connected to the speech act." Donnellan (1974: 17) articulates the point at a bit

more length: “Suppose someone says ‘Socrates was snub-nosed’, and we ask to whom he is referring. ... [T]his calls for a historical explanation; we search not for an individual who might best fit the speaker’s descriptions ... but rather for an individual historically *related* to his use of the name.” Kripke (1972: 94–95) uses similar terms to describe his approach: “... It’s in virtue of our connection with other speakers in the community, going back to the referent himself, that we refer to a certain man ... In general, our reference depends not just on what we think ourselves, but on other people in the community, the history of how the name reached one, and things like that. It is by following such a history that one gets to the reference.” And again Kripke (1972: 106): “... reference actually seems to be determined by the fact that the speaker is a member of a community of speakers who use the name. The name has been passed to him by tradition from link to link.”

The causal-historical theory is an externalist approach to reference, in that reference depends largely on factors external to the speaker’s head – factors pertaining to the speaker’s linguistic community and to the environment in which the expression in question evolved. (Descriptivists tend to be internalists, insofar as they hold that reference is fully determined by the speaker’s beliefs and discriminative abilities.) On the causal-historical view, the criteria for the correct application of a word are not, in general, introspectively accessible to competent speakers; one can competently use ‘gold’ or ‘Aristotle’ without knowing anything that would distinguish Aristotle from Plato, or gold from fool’s gold. Mistaken or ignorant speakers can still single out specific referents via these complex, communal, causal-historical mechanisms. (see **Externalism about Content** for more on this.)

Contra the descriptivists, the causal-historical theorists argue that the meaning of a proper name is not some kind of descriptive sense (see **Direct Reference; Proper Names: Philosophical Aspects; Reference: Philosophical Theories** for discussion). From here, the conclusion that the semantic contribution of a name is just its referent looks compelling. This is why the theory has led to a resurgence of interest in the Millian view of proper names (i.e., the meaning of a name is just its referent) and in the Russellian approach to singular propositions (i.e., the proposition expressed by a sentence containing a name – say, ‘Kaplan is in California’ – is individuated solely in terms of the individual and property that it is about, as opposed to being individuated in terms of more finely-grained concepts or meanings). Many think that the causal-historical chain of transmission story

about how a word refers to something in particular nicely complements, and fleshes out, these doctrines of Mill and Russell.

The causal-historical theory does not aim to give a reductive analysis of reference. For example, Kripke (1972: 96) says: “When the name is ‘passed from link to link,’ the receiver of the name must, I think, intend to use it with the same reference as the man from whom he heard it ... [T]he preceding account hardly *eliminates* the notion of reference; on the contrary, it takes the notion of intending to use the same reference as a given.” (Cf. Kaplan’s [1990] discussion of the point that the intention to preserve reference is not itself a causal notion.) Thus, those who seek to naturalize reference, by reducing the relation of reference to something more scientifically respectable, must either significantly alter the causal-historical view, or look elsewhere.

The Causal Theory of Meaning

In contrast, the causal theory of meaning (also called the ‘information-theoretic’ approach to meaning) is explicitly in the business of explaining semantic phenomena in non-semantic terms. The general aim here is a naturalistic account of the phenomenon of meaning, and the thought is that the notion of causation is the most promising place from which to start. Dretske (1981) is a seminal proponent of this approach, and Fodor (1987, 1990) develops related accounts. Stampe (1977), another influential proponent, gives the following programmatic sketch: “We have causal theories ... of knowledge and memory, of belief, of evidence, of proper and common names, and of reference. If ... these phenomena should turn out to have causal analyses, it will be no mere coincidence. Only their having something in common would make it so ... [The root of this convergence] is that *representation* is essentially a causal phenomenon” (1977: 81).

The general idea behind the causal theory of meaning is that linguistic meaning is a species of causal co-variance. Roughly, the goal is to show that ‘means’ means (more or less) the same thing in (1) and (2), that both cases are, at root, cases of reliable correlation:

1. Smoke means fire.
2. ‘Fire’ means fire.

For a word to mean something in particular is for the word to reliably indicate that thing. Alternatively, a word ‘W’ means M if M tends to cause or bring about tokens of ‘W.’ (The account is intended to apply not only to tokens of ‘W’ that are actually uttered, but also, and more fundamentally, to occurrences of the word in thought.)

If a satisfactory account of meaning were forthcoming down this avenue, this would be a monumental leap forward for the human and cognitive sciences. As yet, there is nothing remotely resembling a satisfactory scientific treatment of meaning; given the fundamental and pervasive roles that meaningful thoughts and utterances play in our lives, that is a rather large gap in our scientific understanding of human beings.

(Note that Grice [1957] criticizes a view that he calls ‘the causal theory of meaning’ – the core of which is the idea that the meaning of an expression ‘E’ is (roughly) the content of the attitude that is prone to cause a speaker to utter ‘E,’ and that hearing ‘E’ is prone to cause in listeners. This view has not played a major role in the philosophy of language; but nonetheless some of Grice’s arguments against it are echoed in the criticisms, described in the next section, of the above information-theoretic causal theory.)

Problems and Prospects

There are many problems with the causal-historical theory of reference (which are discussed at more length in **Reference: Philosophical Theories**). Evans (1973) and Searle (1983) develop counterexamples to the theory, cases where it seems to be committed to unwelcome consequences. Furthermore, many of the semantic views with which the theory has been allied (such as those of Mill and Russell mentioned earlier in the second section of this article) are controversial (*see Direct Reference; Proper Names: Philosophical Aspects* for discussion). More generally, the causal-historical view is just a sketchy picture – it does not offer anything like specific necessary or sufficient causal-historical conditions for identifying the referent of an utterance or inscription. Any utterance stands in an awful lot of causal relations to an indefinite range of things; to single out precisely which subset of these ubiquitous causal relations are semantically relevant – let alone precisely which of them are relevant to determining the referent of a particular use of a particular expression – is a daunting task that is as yet barely begun.

The situation is worse for the (more reductionist, and so more ambitious) causal theory of meaning. It not only falls prey to the problems that befalls the causal-historical approach to reference but also gives rise to some distinctive problems of its own. Basically, for almost any word-meaning pair ‘W’-M, it is not difficult to come up with conditions in which things distinct from M tend to cause ‘W’s, and conditions in which M does not tend to cause ‘W’s. For instance, in various sorts of suboptimal conditions, cows might tend to cause tokens of ‘horse,’ but

nonetheless – regardless of how dark it is (or how far away they are, what they are disguised as, etc.) – these cows are distinct from the meaning of ‘horse.’ In the other direction, if a horse was to ‘baa’ like a sheep, or was painted with zebra-stripes, or what have you, these misleading factors would affect its tendency to cause ‘horse’-tokens, but would not have the slightest effect on the fact that the term ‘horse’ correctly applies to it. In short, causation is a much more indiscriminating relation than meaning; and this is the source of all manner of problems for the project of using causation to build an account of meaning.

There are many refinements of the basic causal-theoretic view, intended to skirt these elementary problems and their many variants. However, the consensus seems to be that this type of causal theory can only succeed in delivering an account of meaning that accommodates our intuitions about the normativity and determinacy of meaning (i.e., respectively, it is possible to misapply a term, and the terms ‘A’ and ‘B’ can differ in meaning even if all As are Bs) if it smuggles in semantic notions, and thus helps itself to meaning, as opposed to offering an account of meaning (for discussion, see Loewer, 1997).

To sum up: the causal theory of reference is the view that a word refers to that to which it stands in the right sort of causal-historical relation. Since the 1970s, it has become the orthodox approach to reference. However, many problems remain to be worked out, for this general picture to yield a satisfactory, comprehensive account of reference. The causal theory of meaning is the view that the meaning of a word is that which reliably causes tokens of the word to be thought or uttered. Many take this to be the most promising avenue for a naturalistic account of meaning. However, there are reasons to think that the approach is too crude to yield an adequate account of linguistic meaning. At best, there are counterexamples that have yet to be satisfactorily addressed.

See also: Direct Reference; Externalism about Content; Proper Names: Philosophical Aspects; Reference: Philosophical Theories; Sense and Reference: Philosophical Aspects.

Bibliography

- Devitt M & Sterenly K (1999). *Language and reality* (2nd edn.). Cambridge, MA: MIT Press.
- Donnellan K (1970). ‘Proper names and identifying descriptions.’ *Synthese* 21, 256–280.
- Donnellan K (1974). ‘Speaking of nothing.’ *Philosophical Review* 83, 3–32.
- Dretske F (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

- Evans G (1973). 'The causal theory of names.' *Proceedings of the Aristotelian Society* 47, 187–208.
- Fodor J (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor J (1990). 'A theory of content' and other essays. Cambridge, MA: MIT Press.
- Grice H P (1957). 'Meaning.' *Philosophical Review* 66, 377–388.
- Kaplan D (1977). 'Demonstratives.' In Almog J, Perry J & Wettstein H (eds.) (1989) *Themes from Kaplan*. Oxford: Oxford University Press. 481–564.
- Kaplan D (1990). 'Words.' *Proceedings of the Aristotelian Society* 64, 93–120.
- Kripkes S (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.

- Loewer B (1997). 'A guide to naturalizing semantics.' In Hale B & Wright C (eds.) *A companion to the philosophy of language*. Oxford: Blackwell. 108–126.
- Putnam H (1975). 'The meaning of "meaning".' In Gunderson K (ed.) *Mind, language, and reality*. Cambridge: Cambridge University Press. 131–193.
- Searle J (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Stalnaker R (1997). 'Reference and necessity.' In Hale B & Wright C (eds.) *A companion to the philosophy of language*. Oxford: Blackwell. 534–553.
- Stampe D (1977). 'Toward a causal theory of linguistic representation.' *Midwest Studies in Philosophy* 2, 42–63.

Character versus Content

C Spencer, Howard University, Washington, DC, USA

© 2006 Elsevier Ltd. All rights reserved.

David Kaplan introduced the content/character distinction in his monograph *Demonstratives* (1989a) to distinguish between two aspects of the meaning of (1) indexical and demonstrative pronouns (e.g., 'I', 'here', 'now', 'this', and 'that') and (2) sentences containing them. Roughly, the content of an occurrence of an indexical or demonstrative is the individual to which it refers, and its character is the rule that determines its referent as a function of context. Thus, an indexical has different contents in different contexts, but its character is the same in all contexts. For instance, the character of 'I' is the rule, or function, that maps a context of utterance to the speaker of that context. This function determines that the content of Sally's utterance of 'I' is Sally.

Content/Character Distinction and Semantics

Sentences containing indexicals or demonstratives are context-dependent in two ways. First, contexts help to determine what these sentences say. Second, contexts determine whether what is said is true or false. For instance, suppose Sally says, 'I'm cold now' at time t . The context supplies Sally as the referent for 'I' and time t as the referent for 'now,' so it helps to determine what Sally said. Other facts about the context, specifically whether Sally is cold at time t , determine whether she said something true or false. Different contexts can play these different roles, as they do when we ask whether what Sally said in one

context would be true in a slightly different context. A central virtue of Kaplan's semantics is that it distinguishes between these two roles of context. For Kaplan, a *context of use* plays the first role, of supplying contents for indexical expressions, and a *circumstance of evaluation* plays the second. A context of use is just a context in which an indexical expression may be used, and which supplies a content for the indexical expression. A circumstance of evaluation is an actual or merely possible situation in which the content of an utterance is evaluated for truth or falsehood.

A semantic framework like Kaplan's, which captures the double-dependence of meaning on context, is sometimes called a two-dimensional semantics. In the two-dimensional framework, a meaningful entity such as a linguistic expression or an utterance determines not a single semantic value but a two-dimensional matrix of semantic values. **Figure 1** represents Kaplan's semantics in this way. In **Figure 1**, the vertical axis of the matrix displays contexts of use (u^1 – u^3) and the horizontal axis displays circumstances of evaluation (c^1 – c^3). Each cell in the matrix gives the extension of the linguistic expression e as used in the specified context of use and evaluated in the specified circumstance of evaluation. In this matrix, the cell in row n and column m gives the

e	c^1	c^2	c^3
u^1	T	F	T
u^2	F	F	F
u^3	T	T	T

Figure 1 Two-dimensional matrix.

semantic value of e in the context of use specified at the beginning of row n and evaluated in the circumstance of evaluation specified at the top of column m . If e is a sentence, cells will be filled in with truth values as illustrated.

Kaplan offers a syntax and semantics for a formal language containing indexicals, demonstratives, and a variety of modal operators. In this formal system, a context of use is an ordered n -tuple of contextual features to which indexicals or demonstratives are sensitive, such as the speaker, time, world, and location of the context. A circumstance of evaluation is an ordered n -tuple of a possible world-state or world-history, a time, and perhaps other elements as would be required given the sentential operators in the language. For Kaplan, all contexts of use are *proper*, which means that the speaker of the context must be located at the time, place, and world of the context. Circumstances of evaluation, however, need not be proper.

Contexts of use and circumstances of evaluation play a role in the specification of the character and content of an expression. The character of any linguistic expression e is a function from contexts of use to contents appropriate for e , i.e., an individual if e is a singular term, a proposition if e is a sentence, and sets of n -tuples of individuals if e is an n -place predicate. Indexical expressions only have contents relative to a context of use. So Kaplan speaks of the content of an occurrence of an expression rather than the content of the expression itself. Contents are evaluated in circumstances of evaluation, and these evaluations yield extensions appropriate to the kind of content under evaluation. So we also can characterize the content of an occurrence of e as a function from circumstances of evaluation to extensions of a type appropriate to e . For instance, the extensions for sentences are truth values, for indexicals, individuals, and for n -place predicates, n -tuples of individuals. For individuals and n -place predicates, these will be constant functions (i.e. the function delivers the same extension in every circumstance of evaluation). It is often simpler to think of contents as individuals (for singular terms), propositions (for sentences) and sets of n -tuples of individuals (for n -place predicates), and Kaplan typically talks about contents in this way. Both ways of thinking of contents are semantically equivalent.

For Kaplan, indexicals and demonstratives are both directly referential and rigidly designating. They are directly referential because they contribute only their referents to the propositions expressed by sentences containing them. They are rigidly designating because, once they secure a content in a context of use, they retain that content in every circumstance of evaluation. Indexicals and demonstratives contrast

with the typical definite description in both respects. Definite descriptions typically contribute a descriptive condition to a proposition rather than an individual, and this descriptive condition is typically satisfied by different individuals in different worlds of evaluation. Although Kaplan's view that demonstratives are directly referential is widely accepted, some recent discussions of complex demonstratives (i.e. expressions of the form 'that F ') have defended a quantificational approach, and some considerations in favor of such an approach may apply to the pure demonstratives 'this' and 'that' (King, 2001).

Kaplan's semantics has technical virtues lacking in earlier treatments of natural language indexicality. It shares with other double-indexing accounts (Kamp, 1971) a technical superiority to single-index theories, which evaluate sentences relative to a single index, which is an ordered n -tuple of features of a context, such as a speaker, time, location, and world. Such theories cannot account for the interaction of indexicals and certain sentence operators. To evaluate the sentence (1), for instance, we need to consider the truth value of the constituent sentence, 'the man who is now President of the United States no longer hold[s] that office' in situations occurring *after* the sentence is uttered.

- (1) Someday, the man who is now President of the United States will no longer hold that office.

But the indexical 'now' in that constituent sentence must still refer to the time (1) is used, and not the time at which the constituent sentence is evaluated. As Hans Kamp has argued, only a double-indexing theory will correctly predict the truth conditions for (1) (Kamp, 1971).

Content/Character Distinction and Philosophy

The content/character distinction sheds light on some specifically philosophical issues involving context-sensitivity in thought and language. These applications involve philosophically significant assumptions, and are more controversial than the applications to the semantics of indexicals and demonstratives.

First, content and character play two roles that Gottlob Frege initially envisioned for the meaning, or sense, of a sentence, one semantic and the other more broadly psychological (Frege, 1892). Frege thought that the sense of a sentence should both determine its truth condition and provide the cognitive significance of beliefs expressible with that sentence. Although Frege expected that one entity, the sense, could play both roles, indexical and demonstrative belief undermines this expectation, since it

appears to require two different entities to play the two roles. Different people who have a belief they could express by saying 'I'm cold' will be in the same psychological/functional state. They will all be shivering and trying to get warmer. But because each person who thinks, 'I'm cold' is a constituent of the content of that thought, all of these thoughts will differ in content. The psychological role of an indexical belief appears to be more closely tied to the character of the sentence the thinker would use to express that belief than to the content of the belief. But the content, rather than the character, is more directly relevant to the truth condition of an occurrence of a sentence containing an indexical.

Second, Kaplan has suggested that the content/character distinction helps to explain the relation between the epistemological notions of logical truth and the *a priori*, on the one hand, and the metaphysical notions of necessity and contingency on the other. Other philosophers have put broadly similar applications of the two-dimensional framework into service to the same end (Stalnaker, 1978, cf. Stalnaker, 2004; Chalmers, 1996; Jackson, 1998). As is evident to anyone who understands sentence (2), it cannot be uttered falsely. Therefore, sentence (2) is in a certain sense a logical or *a priori* truth. Yet it does not express a necessary truth, since occurrences of (3) will typically be false.

(2) I am here now.

(3) Necessarily, I am here now.

Kaplan has suggested that we explain the special status of (2) as follows: metaphysically speaking, (2) is contingently true in virtue of its content. But it has its special epistemic status in virtue of its character: the character of (2) requires that it express a truth in every context of use. Other sentences which may express the same content as a particular occurrence of (2), but a different character, such as (4), do not have the same special epistemic status.

(4) GWB is in Washington, DC on June 16, 2004.

Because (4) and some occurrences of (2) share a content but differ in their epistemic status, it is natural to conclude that contents cannot be the bearers of this special epistemic property. Critics of this account of the *a priori* (Soames, 2005) say that the content/character distinction cannot underwrite the general account of *a priori* knowledge that some of its defenders (Chalmers, 1996; Jackson, 1998) have claimed.

Third, some philosophers have used Kaplan's content/character distinction to distinguish narrow content (i.e. content determined by the internal state of the thinker) from wide content (i.e. content determined by the internal state of the thinker and his or

her environment) (Fodor, 1987; see also Chalmers, 1996; Jackson, 1998, for a related application of two-dimensional semantics to these ends). They suggest that narrow content is loosely modeled on Kaplan's characters, and wide content on Kaplan's contents. That characters seem to capture something important about the psychological roles of belief makes them particularly attractive candidates to model the purely internal aspects of thought. Critics of the approach contend that although characters help to characterize internal states of thinkers, they are not themselves determined by such states (Stalnaker, 1989).

See also: Analytic/Synthetic, Necessary/Contingent, and *a Priori/a Posteriori*; Distinction; Direct Reference; Essential Indexical; Indexicality: Philosophical Aspects; Reference: Philosophical Theories; Rigid Designation; Situation Semantics.

Bibliography

- Almog J, Perry J & Wettstein H (1989). *Themes from Kaplan*. New York: Oxford University Press.
- Chalmers D (1996). *The conscious mind*. New York: Oxford University Press.
- Fodor J A (1987). *Psychosemantics: the problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Frege G (1892). 'Ueber sinn und bedeutung.' *Zeitschr. fur Philos. und Philos. Kritik* 100. Feigl H (trans.). 190–202.
- Jackson F (1998). *From metaphysics to ethics*. New York: Oxford University Press.
- Kamp H (1971). 'Formal properties of "now".' *Theoria* 37, 227–273.
- Kaplan D (1989a). 'Demonstratives.' In Almog, Perry, & Wettstein (eds.). 481–564.
- Kaplan D (1989b). 'Afterthoughts.' In Almog, Perry, & Wettstein (eds.). 565–614.
- King J (2001). *Complex demonstratives*. Cambridge, MA: MIT Press.
- Kripke S (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Lewis D K (1980). 'Index, context, and content.' In Kanger S & Ohman S (eds.) *Philosophy and grammar*. Dordrecht: Reidel.
- Soames S (2005). *Reference and description: the case against two-dimensionalism*. Princeton, NJ: Princeton University Press.
- Stalnaker R C (1978). 'Assertion.' In Cole P (ed.) *Syntax and semantics, vol. 9: pragmatics*. New York: Academic Press, Inc. 315–322.
- Stalnaker R C (1989). 'On what's in the head.' *Philosophical Perspectives* 3, *Philosophy of Mind and Action Theory*. 287–316.
- Stalnaker R C (2004). 'Assertion revisited: on the interpretation of two-dimensional modal semantics.' *Philosophical Studies* 118(1–2), 299–322.

Cognitive Science and Philosophy of Language

S Scott, Washington University in St. Louis,
St. Louis, MO, USA

© 2006 Elsevier Ltd. All rights reserved.

Much contemporary philosophy of language can be viewed as a synthesis of three major traditions: ideal language philosophy, ordinary language philosophy, and cognitivism. In the first three-quarters of the 20th century, philosophers in both the ordinary and ideal language traditions sought to solve or dissolve traditional philosophical problems through careful exegesis of the meanings of words and sentences. For ideal language philosophers, the project was to formally describe how words and sentences ought to be interpreted in scientific and philosophical discourse. For ordinary language philosophers, the project was to characterize the conventions underlying the actual use of words and sentences in ordinary speech. Philosophers in both traditions made a number of lasting contributions to the philosophical and scientific study of language, but they were not just studying language for its own sake. Many philosophers in this period considered the philosophy of language to be first philosophy, the foundation on which other philosophical inquiries are built, and they had other philosophical issues in mind when developing their accounts of language (*see Epistemology and Language; Metaphysics, Substitution *Salva Veritate* and the Slingshot Argument*).

As the limitations of the ordinary and ideal language traditions became apparent and their influence began to decline, the cognitivist tradition in the scientific study of language was growing. Cognitivists view the mind as a computational and representational system and bring a wide variety of empirical evidence to bear on their investigations into the structure and processing of linguistic knowledge in the mind. The synthesis of cognitive science and philosophy of language, or as I shall call it, the new philosophy of language, integrates the formalisms of the ideal language tradition with the careful attention to the nuances of use that characterized the ordinary language tradition. But as cognitivists, many contemporary philosophers of language also take results from linguistics into account and share with other cognitive scientists a commitment to producing theories that are consistent with available psychological and neuroscientific evidence. What follows is a very brief account of the three traditions and their synthesis into the new philosophy of language, ending with a review of some recent work on proper names that exemplifies this new synthesis.

The Ideal Language Tradition

Ordinary speech is a rich source of vagueness, ambiguity, puzzles, and paradoxes, most of which go unnoticed by most speakers. This may not matter all that much for the purposes of ordinary conversation, but in scientific and philosophical discourse the imprecision of ordinary language is not to be tolerated. So said Bertrand Russell, Gottlob Frege, W. V. O. Quine, and the philosophers of the ideal language tradition. According to them, ordinary language contains certain deficiencies and the philosopher's job is to provide remedies (Russell, 1919: 172, describes one such "deficiency" as a "disgrace to the human race"). The goal of these philosophers was to standardize and regiment language, explain away puzzles and paradoxes, and formally characterize ambiguities. Their aim was to transform ordinary language into something closer to an ideal language – one that scientists and philosophers could use to express their hypotheses about the world. The strengths and weaknesses of their approach can be illustrated using Russell's theory of proper names.

Example: Proper Names

The idea that scientific hypotheses are about the world was key for the ideal language philosophers. Sentences in science and philosophy, not to mention ordinary conversation, often attribute properties to objects in the real world (*see Objects, Properties, and Functions*). Accordingly, a defining feature of ideal language philosophy was the idea that the relationship of reference is a basic unit of meaning (*see Reference: Philosophical Theories*), and the starting point was the analysis of simple property attribution sentences such as:

- (1a) Venus is round.
- (1b) Venus is a star.

Here are some basic intuitions: Sentence (1a) is true because the planet Venus has the property of being round, and sentence (1b) is false because the planet Venus does not have the property of being a star.

Here is a simple analysis that respects those intuitions: In both sentences, the proper name *Venus* refers to an object (*see Proper Names: Philosophical Aspects*), the remaining words *is round* and *is a star* attribute properties to that object, and the sentences refer to the propositions that Venus is round and that Venus is a star, respectively (*see Propositions*). This analysis is shown more formally in (2), where VENUS denotes the actual object Venus, not a word or an idea.

Simple Analysis of (1)

(2a) *round*(VENUS)

(2b) *star*(VENUS)

This analysis of simple sentences can be developed into a powerful system for characterizing the semantics of much more complex and interesting sentences. But, unfortunately, it also runs into fatal problems with certain sentences that seem just as simple as those in (1). For instance, it is not easy to see how to extend the analysis to cover:

(3) Vulcan is round.

This sentence was once thought to be true by astronomers who postulated the existence of a planet, tentatively named *Vulcan*, to explain the observed perturbations in Mercury's orbit. It is now known that there is no such planet or, to put it another way, that *Vulcan* is an empty name (see **Empty Names**). So, although (3) is clearly meaningful and has a grammatical form that parallels the sentences in (1), the simple analysis will not work in this case. Recall that (1a) is true because the object referred to by the name *Venus* has the property of roundness. But in (3), there is no object named *Vulcan* and therefore nothing to which any such property can be applied.

Here we have the makings of a puzzle – if reference is as basic to meaning as it appears to be, then how is it possible to say meaningful things using words that have no referents? One option is to allow that nonexistent things such as Vulcan, Santa Claus, unicorns, and so on really do have some kind of objecthood. But most philosophers would reject this option because, as Russell (1919: 169) put it, “logic . . . must no more admit a unicorn than zoology can; for logic is concerned with the real world just as truly as zoology.” Another option is to just bite the bullet and accept that (3) does not express a proposition and is therefore meaningless. Although some contemporary philosophers of language have taken this route (e.g., Adams and Stecker, 1994), the ideal language philosophers did not want to take that way out either because to do so would be to render many important scientific and philosophical hypotheses meaningless.

Russell's Theory of Descriptions

Russell found a solution to the problem of empty names (and other philosophical puzzles) in his theory of descriptions (see **Descriptions, Definite and Indefinite: Philosophical Aspects**). Briefly, Russell held that names such as *Vulcan* and *Venus* do not directly refer but instead are shorthand for definite descriptions such as *the planet causing perturbations in*

Mercury's orbit and *the second planet from the sun*, respectively. That is, names are disguised definite descriptions. So, when scientists utter sentences such as those in (1) and (3), what they assert is something more like:

Step One of Russell's Analysis of (1) and (3)

(4a) The second planet from the sun is round.

(4b) The second planet from the sun is a star.

(4c) The planet causing perturbations in Mercury's orbit is round.

On the face of it, it looks like (4c) has the same problem as (3) – descriptions such as *The planet causing perturbations in Mercury's orbit* seem like they should be interpreted as namelike referring expressions. But Russell did not think so. He thought that descriptions such as these should be analyzed as general, quantificational statements about what exists in the world. In the case of (4c), the correct interpretation, according to Russell, is that there is exactly one planet causing perturbations in Mercury's orbit and all such planets are round. This analysis is expressed in quantificational notation in (5), where *pm*() stands for the property of being a planet that causes perturbations in Mercury's orbit. (Some of the inessential details have been simplified in a way that Russell might have objected to, but that does not matter for current purposes.)

Step Two of Russell's Analysis of (3)

(5a) There exists exactly one planet *x* that is the cause of the perturbations in Mercury's orbit, and it is round.

(5b) $\exists x ((\forall y (pm(y) \leftrightarrow y = x)) \ \& \ round(x))$

In this final analysis, there is no longer any element in the proposition corresponding to the name *Vulcan* and no role available for any referent, and thus the puzzle of empty names disappears. To recap: Names are shorthand for disguised definite descriptions, and sentences that contain definite descriptions express general propositions about the world and the things in it rather than singular propositions about particular entities.

Limitations of the Ideal Language Approach

Russell's analysis of proper names, as clever and influential as it is, runs afoul of ordinary intuitions. Sentence (3) seems to have a very simple subject-predicate form, but the proposition in (5) that provides the meaning for (3) bears no resemblance to that form. Furthermore, (5) is false because it asserts the existence of something that does not exist (i.e., it asserts the existence of a planet that causes perturbations in Mercury's orbit, but there is no such planet). But it is not clear to everybody that (3) really is false

(see Strawson, 1950; and the reply by Russell, 1957). To many people, questions such as *Is Vulcan round?* have the same kind of problem as questions such as *Have you stopped cheating on exams yet?* – to answer either “yes” or “no” would be to accept a problematic premise.

Russell was not driven to this analysis of simple sentences as an attempt to characterize how ordinary speech works but as an attempt to dissolve an apparent logico-scientific puzzle that arises when we take the referential commitments of ordinary speech seriously. But the analysis ends up providing no account of the fact that people seem quite capable of making what appear to be true claims about nonexistent things.

- (6a) Santa Claus usually wears a red suit.
- (6b) Pegasus looks a lot like a horse.

Russell’s theory of disguised definite descriptions makes the sentences in (6) come out false, contrary to most people’s intuitions. His theory preserves the apparent meaningfulness of these sentences, and does so without maintaining any problematic commitments to entities such as Pegasus and Santa Claus, but at the price of a theory that may not have much to say about their ordinary use.

The Ordinary Language Tradition

As vague, ambiguous, and rife with semantic puzzles as ordinary language is, it also contains a wealth of information that philosophers cannot afford to ignore. In order to discover anything meaningful about important philosophical topics such as Truth, Knowledge, and Justice, philosophers need to know what *truth*, *knowledge*, *justice*, and other related words actually mean in ordinary language. This was the perspective of Gilbert Ryle, H. P. Grice, J. L. Austin, P. F. Strawson, Ludwig Wittgenstein (in his later works), and the philosophers of the ordinary language tradition. According to them, philosophers must pay careful attention to the nuances of ordinary language use and must be particularly wary of misusing ordinary language expressions in their philosophical theories. In many ways, this tradition was radically opposed to the ideal language tradition: whereas the ideal language project was a prescriptive project, concerned with legislating how language ought to be understood, the ordinary language approach was purely descriptive, concerned with the investigation of how language is actually used; whereas ideal language philosophers sought to construct a theory of meaning based on reference to things in the world, ordinary language philosophers sought to construct a theory of meaning

based on conventions of ordinary use (see *Use Theories of Meaning*). But despite these differences, both traditions shared a common motivation, namely, the analysis of language in order to help solve or dissolve philosophical problems. It is just that in pursuing this common aim, ideal language philosophers were busy constructing a new language while ordinary language philosophers were busy pointing out how philosophers tended to misuse the old one.

Example: Ryle on Free Will

Ordinary language philosophers thought that the meaning of an expression is the conventions governing its use. Thus, to get at the meaning of an expression, we have to examine how it is ordinarily used. The standard technique is to assemble a list of sentences containing a given expression and then try to find conditions under which it would be inappropriate or nonsensical to use those sentences. Whatever those conditions turn out to be, their negation must be part of the meaning of the word in question. (Notice that this makes short work of the puzzle of empty names. Because the meaning of a word is the conventions governing its use, names can have meaning whether they have a referent or not.)

As an example of ordinary language analysis in action, consider Ryle’s (1949) investigation of the word *voluntary*. Ryle noted that philosophers often characterize free will (another important philosophical topic) based on the distinction between voluntary and involuntary actions – free will is said to be involved in an action when it is performed voluntarily and not when it is performed involuntarily. So *voluntary* (along with grammatical variants such as *voluntarily*) is an important word in philosophy, but what does it actually mean in ordinary language? Consider the following sentences:

- (7a) Benazir went to school voluntarily.
- (7b) Hussein ate the sandwich voluntarily.
- (7c) Ahmad watched *Seinfeld* voluntarily.

As Ryle observed, such uses of *voluntary* and its grammatical variants seem odd or wrong in any situation in which there is no reason to believe that the person in question ought not to have performed the action. So if Benazir has been banned from campus or hates school or is supposed to be doing something else, then (7a) might make sense. But if there is no reason to suppose anything like that, then the word *voluntarily* should be left out. Ditto for (7b) and (7c). From these sorts of considerations, Ryle concluded, part of the meaning of the word *voluntary* must include the condition that it can only be used in the description of an action that for some reason ought not to have been performed.

To bring this back to the philosophical problem of free will, Ryle noted that philosophers who worry about what it could mean to eat a sandwich or watch *Seinfeld* voluntarily, absent any kind of context, are systematically misusing ordinary English. As he put it, they engage in an “unwitting extension of the ordinary sense of ‘voluntary’ and ‘involuntary’ ” (Ryle, 1949: 69). The conclusion that Ryle drew from these and other considerations was that there is no problem of free will. The appearance of the problem arises only when philosophers misuse ordinary language words such as *voluntary*. The whole problem just dissolves under ordinary language analysis.

Limitations of the Ordinary Language Approach

The ordinary language philosophers tended to be less likely to make use of formalisms for characterizing the meanings of words or sentences. The nature of ordinary language analysis was such that it produced accounts of word or sentence meaning that tended to be less rigorous than those produced by philosophers working in the ideal language tradition. Furthermore, the use theories of meaning pursued by ordinary language philosophers had little to say about the relationship between language and reality, and were thereby limited in their ability to account for reference and truth conditions, whether in scientific, philosophical, or ordinary discourse. The ordinary language philosophers demonstrated many of the important and subtle ways in which philosophically interesting words are employed in ordinary language, but they did so at the price of having neither a systematic, precise account of meaning nor a theory of the relationship between language and the world.

The ordinary language tradition ultimately met its demise at the hands of its own adherents. In his 1967 lectures on ‘Logic and Conversation,’ Grice (1989) gave a strong voice to many philosophers’ growing misgivings about the project. He argued for a sharp distinction between what is said by a speaker on a particular occasion and what the speaker might have meant by what was said. For Grice, what is said is the literal, truth-evaluable, relatively invariant portion of meaning. To use one of his examples, suppose Alyssa happens upon Cliff, who has run out of gas on the highway, and utters:

(8) There’s a gas station around the corner.

What Alyssa has said, in Grice’s sense (literally expressed, truth-conditional meaning) is the proposition that around the indicated corner is a gas station. Alyssa said nothing further about whether the gas station is open, has gas to sell, and so on. But assuming she is sincerely trying to help Cliff out, it will be inappropriate for her to use that sentence unless she

believes that the gas station is open and has gas to sell. Based on this latter observation, an ordinary language philosopher might be tempted to conclude that these further conditions are part of the meaning of (8). But that, Grice argues, is a mistake.

Grice’s alternative is that the further propositional content about the gas station being open and having gas to sell is not part of the literal meaning of (8), but is what he called a conversational implicature. This conversational implicature is part of what Alyssa means to communicate with (8), but she expects Cliff to be able to pick up on it without requiring her to state it explicitly. The details of how Cliff might do that is beyond the scope of the current discussion (see Grice, 1989; Sperber and Wilson, 1995), but to get a sense of the reasonableness of the distinction between what is said and what is conversationally implicated, consider how Alyssa could have tacked an extra clause onto (8) to take back both what she implicated and what she said.

Clauses That Cancel Implicatures

- (9a) There’s a gas station around the corner, but it’s not open.
- (9b) There’s a gas station around the corner, but it’s out of gas.

The sentences in (9) both have (8) embedded in them, and the fact that they do not seem contradictory indicates that the material in the final clause must not be opposed to any part of the meaning of (8). Now suppose Alyssa had instead uttered one of the sentences in (10).

Clauses That Contradict What Is Said

- (10a) There’s a gas station around the corner, but it’s not a gas station.
- (10b) There’s a gas station around the corner, but it’s not around the corner.

The fact that these sentences are clearly contradictory indicates that the added clauses must be opposed to some part of the literal meaning of (8). So there is strong intuitive support for the distinction between what Alyssa has said, as shown by the contradictory clauses in (10), and what she conversationally has implicated, as shown by the noncontradictory clauses in (9).

On the basis of this distinction, Grice argued for caution when moving from facts about how words are used to facts about the meanings of those words. It would have been inappropriate for Alyssa to utter (8) if she thought the gas station was closed, but that does not tell us anything about what (8) means. Evidence about use can, in principle, indicate something about the literal meaning of words and sentences, but not always in such a simple way. Ryle, in particular, was

probably wrong to jump from facts about the use of the word *voluntary* to facts about its meaning (and then to the denial of the problem of free will). Grice thought that ordinary language analysis could still be useful but that philosophers needed to pay more attention to separating what an expression can be used to communicate from what that expression actually means in the language – a project that turns out to be exceedingly difficult (*see Semantics–Pragmatics Boundary*).

The Cognitivist Tradition

Language is a fascinating topic of study in its own right, regardless of its role in helping philosophers do their work. It is now clear that the production of even very simple speech behaviors is far more complex than was once thought and working out how linguistic knowledge is structured and processed in the human mind should be a central goal in the scientific study of language. That is what linguists working in the cognitivist tradition tend to think. According to them, the goal of linguistic inquiry is not primarily to account for reference and truth or to characterize conventions of use but rather to find out what it is about the human mind that makes language what it is. Cognitivism is actually a cross-disciplinary tradition concerned with the study of the human mind in general, not just language. Leading figures in the birth and early development of the cognitivist tradition included computer scientists (e.g., Marvin Minsky; psychologists (e.g., George Miller), linguists (e.g., Noam Chomsky, and philosophers (e.g., Hilary Putnam, Jerry Fodor, Daniel Dennett).

There are four features that, taken together, loosely define the cognitivist approach to the study of mind and language: (1) an adherence to computational and representational theories of mind, (2) a rejection of most forms of behaviorism, (3) an openness to empirical evidence from a wide variety of sources, and (4) a tendency toward identifying linguistic meanings with mental states rather than with things in the world or patterns of ordinary use. Each of these aspects is discussed next.

Computational and Representational Theories of Mind

Cognitivists model the mind/brain as an information processing system that performs computations on structured representations of the world. In other words, the mind/brain is a kind of computer, analogous in many ways to a digital computer. Many people find this claim jarring at first, but actually it is quite natural to suppose that, at least in some circumstances, people use computers to do their thinking for

them. Whenever an accountant uses a spreadsheet to prepare tax forms, a pilot flies using an automatic guidance system, or a librarian searches an electronic catalog, computers are being used to perform tasks that would require mental effort if performed by human beings. When people use a computer to perform a task, they avoid some of the thinking that would have been required if they had performed the task unaided.

Digital computers accomplish their apparently mental feats by executing algorithms that manipulate data structures. An algorithm is a set of fully explicit, step-by-step instructions for accomplishing a given task, and a data structure is a package of information about some aspect of the world. For example, a data structure might contain information about a social hierarchy, the layout of a city, or the structure of a sentence. Algorithms contain instructions for how to use those data structures to decide, for example, who to approach for a loan, how to get from downtown to the suburbs, or what a speaker might mean by uttering a particular sentence. Cognitivists claim that human thought consists of computational processes (analogous to algorithms) that operate on mental representations of the external world (analogous to data structures), although there remains much debate over the nature of those processes and representations.

Like a digital computer, the mind/brain can be analyzed at a number of different levels (Dawson, 1998; Marr, 1982). At the physical level, digital computers are instantiated in electronic circuitry and minds are instantiated in brains. By investigating the brain, we can figure out what kinds of mental representations and computational processes it supports and what parts of it may or may not be involved in language. At the algorithmic level, digital computers run programs that specify the details of their behavior. The bold conjecture of cognitive science is that minds are the programs that run on the physical circuitry of the brain. By performing psychological experiments, we can shed light on how linguistic knowledge is represented in the mind and what computational processes are involved in using that knowledge. Finally, there is the task level. The programs that digital computers run can only be made sense of in light of knowledge about their connections to the world and the tasks they were designed to solve. Similarly, in order to understand how the mind uses language, it is necessary to have a theory of what language is and what knowledge is involved in language use.

These three levels of analysis thus define a multi-disciplinary program of research into the nature of human language, with different research questions

posed at each level (see Table 1). Cognitivist linguists focus most of their attention on the algorithmic and task levels, concentrating on the difficult problems of identifying the knowledge required to produce well-formed grammatical utterances, determining how that knowledge must be represented in the minds of the speakers, and identifying which elements of that knowledge are learned and which are innate (see **Innate Knowledge**). But as cognitivists, they remain open to, and sometimes make use of, evidence from the physical level as well.

The Rejection of Linguistic Behaviorism

Prior to the establishment of the cognitivist tradition in the 1960s and 1970s, the dominant approach to the study of the mind and language was behaviorism. Many philosophers at the time endorsed or were influenced this approach, including prominent representatives of both the ideal language and ordinary language traditions. Behaviorism comes in a number of varieties (see **Behaviorism: Varieties**), but what all behaviorists agree on is a rejection of internal mental states as something that can be scientifically studied or appealed to in explanations of language and behavior. For psychologists such as B. F. Skinner, this meant that linguistic behavior was to be explained as a complex pattern of responses to environmental stimuli. Verbal responses were thought of as being under the control of certain stimuli in the environment (Skinner, 1957).

Skinner's view of language was subjected to ruthless criticism from Chomsky, who pointed out the complexity of linguistic behavior and the wide variety of possible responses to a given stimulus:

A typical example of stimulus control for Skinner would be the response to ... a painting with [the utterance] *Dutch*. ... Suppose instead of saying *Dutch* we had said, *Clashes with the wallpaper, I thought you liked abstract work, Never saw it before, Tilted, Hanging too low, Beautiful, Hideous, Remember our camping trip last summer?* (Chomsky, 1959, p. 31)

Table 1 Three-level research program

Level	Questions
Task	How are natural languages structured? What must people know and what must they know how to do in order to produce and understand human speech?
Algorithmic	How is knowledge of language represented in the mind? What computational processes are involved in producing and understanding speech?
Physical	How are these representations and computational processes implemented in the hardware of the brain?

Once the nonstimulus-bound nature of linguistic behavior is fully appreciated, said Chomsky, the prospect of arriving at an account of linguistic behavior without involving an appeal to mental states is completely hopeless. Cognitivism pointed the way out of behaviorism by providing a method of formally characterizing those mental states.

The Open Evidence Base

The cognitivist tradition is an empirical tradition. The sources of evidence available to the linguist include the judgments of native speakers, the process of first-language acquisition, the controlled psychological study of speech production and comprehension, the study of acquired and genetic language deficits, and the study of the neurological features of language use in healthy adults, to name but a few. These sources of evidence can be used to investigate language at the task, algorithmic, and physical levels (see Table 2). This is not to say that it is the current practice of linguists to make use of all of these sources of evidence. Indeed much work in theoretical linguistics proceeds using only the grammaticality judgments of the linguists themselves. But there is a general commitment both to the idea that a complete theory of language has to be consistent with all these sources of evidence and to the idea that the evidence base for linguistics is open – that is, there are no principled limits on the kinds of evidence that might bear on the structure of linguistic knowledge.

The commitment to an open evidence base has important consequences. For behaviorists, the study

Table 2 Sources of evidence for the three levels

Level	Example sources of evidence
Task	Judgments of native speakers Which strings of words are grammatical and which are not? What meanings can a sentence have and not have?
Algorithmic	Developmental psychology How do children acquire language? What are the common patterns of language development? Cognitive psychology How do adults react to linguistic stimuli under controlled conditions?
Physical	Clinical studies What kinds of brain injuries and diseases cause language deficits? What specific language deficits are caused by specific brain injuries and diseases? Anatomical and functional studies What parts of the brain are involved in language use? How are these parts interconnected?

of language had to be grounded in observable behavior only. As Quine (1960) pointed out, it turns out that this leads to the conclusion that linguistic knowledge and meaning cannot be unambiguously determined. From this, he drew the conclusion that there is simply no fact of the matter about how to specify the mapping from words and sentences to their meanings (see **Indeterminacy, Semantic**). A famous response to Quine, again from Chomsky (1969), is based on the notion of the open evidence base. According to Chomsky, Quine reached his radical conclusions about semantic indeterminacy by accepting in advance the behaviorist notion that only observable behavior and responses to environmental stimuli may be used as the data for theories of linguistic meaning. But, as Chomsky points out, no other science places such *a priori* limits on the kinds of evidence that can be used to decide between competing theories. As long as the evidence base in linguistics remains open, the possibility of discovering further evidence that will help determine linguistic meaning is open as well.

Meanings as Mental States

The establishment of a viable theory about mental states and mental processing opened the door to a new class of theories of linguistic meaning based on the pairing of words in the public language with mental states of speakers. The general idea of a mental state theory of meaning is at least as old as Aristotle (see **Aristotle and Linguistics**), but the computational and representational theory of mind gave it new life by providing a story about what mental states might be like and how they might be processed in the mind. In addition to endorsing a mental state account of meaning, some cognitivists also harbor a deep mistrust of the reference-based theories pursued in the ideal language tradition. The semanticist Ray Jackendoff (2002) argues that the only kind of reference a cognitivist theory of language can countenance is reference to other mental states, whereas Chomsky (2000) suggests that reference, as originally construed by ideal language philosophers, is not a suitable topic for scientific inquiry at all.

Jerry Fodor (1975) has proposed that words and sentences come by their meaning through being paired with internally represented formulae in what he calls the Language of Thought, or Mentalese (see **Mentalese**). Mentalese is not a public language such as English. It is more like a computer language – a formal system with a combinatorial syntax and an expressive power that equals or surpasses that of a public language. Fodor proposes that words and sentences express mental states, but, unlike

Chomsky and Jackendoff, he takes the further step of attempting to scientifically characterize the meanings of expressions in Mentalese as relationships to objects and properties in the external world (see **Representation in Language and Mind; Causal Theories of Reference and Meaning**). Fodor's theory of meaning thus has two parts: (1) words inherit their meanings from the mental states they express, and (2) most of those mental states get their meanings through reference to the external world. An important alternative cognitivist account of meaning as mental states is offered by connectionism, although a full discussion of that approach is beyond the scope of this article.

The Limitations of Cognitive Science

It is not yet clear how far cognitive science can go, and there are philosophers who dispute the claim that studying the structure and processing of linguistic knowledge in the human mind can tell us much about the nature of language itself (see Barber, 2003). But the computational and representational theory of mind, as a working hypothesis, has given rise to a productive research program producing theories of mind and language rich and predictive enough that, at the very least, they should not be ignored. The cognitivist approach to the study of mind and language is widely regarded by philosophers as the only approach currently worth taking seriously.

The New Philosophy of Language

The new philosophy of language emerged in the 1970s as a synthesis of the ideal language, ordinary language, and cognitivist traditions. From the ideal language tradition comes the use of rigorous formalisms and a concern for the connection between language and reality. From the ordinary language tradition comes the descriptive nature of the project and careful attention to the nuances of ordinary use, as well as Grice's distinction between what is said and what is implicated by an utterance. And from the cognitivist tradition comes an adherence to computational and representational theories of the mind, a rejection of linguistic behaviorism, an attention to the mental states of the language user, and a concern with making semantic and pragmatic theories consistent with the relevant empirical results concerning language and the mind.

The boundaries between linguistics and the philosophy of language have become blurred in this new synthesis. Whereas phonology (the sounds of language), morphology (the structure of words), and syntax (the structure of sentences) remain a concern

mostly of linguists, semantics (the meaning of language) and pragmatics (the communicative use of language) are studied by both linguists and philosophers. There has also been considerable cross-fertilization between linguistics and philosophy. Linguists have adopted the formalisms of the ideal language tradition and the Gricean view of the relation between semantics and pragmatics that arose out of the ordinary language tradition. Philosophers, on the other hand, have adopted the linguistic account of syntax and feel an obligation to relate the semantic interpretation of a sentence to its syntactic form. In addition, the cognitivist approach to linguistics also throws up a host of difficult conceptual issues that demand a rigorous philosophical treatment (*see Philosophy of Linguistics*), for example, the place of reference in semantic theory (*see Externalism about Content*), the nature of linguistic knowledge (*see Innate Knowledge; Tacit Knowledge*), and the connection between language and thought (*see Thought and Language: Philosophical Aspects*).

Two More Theories of Proper Names

How might a practitioner of the new philosophy of language tackle a traditional semantic problem such as the content of proper names? Two theories of proper names are presented by Tyler Burge (1973) and Larson and Segal (1995). These two theories agree with one another in many important respects – so much so that we might be tempted to suppose that they are merely variants of one another. But, as Gabriel Segal (2001) points out, there are a number of pieces of relevant evidence from the task, algorithmic, and physical levels of cognitive analysis that may be used to adjudicate between the theories. (A caution: The semantic issue is actually more technical than the following discussion suggests, concerning points of difference between semanticists working in the formal framework of truth-theoretic semantics. Because there is no room to introduce the details of that framework here, the accounts of the rival theories are somewhat sketchy, although, I hope, detailed enough to make it clear how empirical evidence can be used to decide between them.)

Burge's approach to proper names is a variation on Russell's disguised definite descriptions. Burge proposes that proper names are actually a kind of common noun, that is, words such as *table* and *cat* that encode properties that apply to large numbers of objects. In Burge's account, if we have a cat named *Sylvester*, then that object has both the property of being a cat (a property it shares with other cats) and the property of being a Sylvester (a property it shares with other Sylvesters). In defense of this idea, Burge

points out that, like common nouns, names can be pluralized and paired with determiners such as *the* and *a*:

- (11a) There are very few Sylvesters in the world.
- (11b) There were three Madelines at the party.
- (11c) There's a Bartholomew Kropotnik here to see you.
- (11d) The Jessica I met today was a real jerk.

This idea encounters an immediate difficulty. Burge says that names are common nouns, even when they occur unmodified and on their own:

- (12) Fido wants to chase Sylvester.

But other common nouns cannot be used that way in English:

- (13) *Dog wants to chase cat.

Sentence (13) only works if we interpret *dog* and *cat* as unusual names rather than as common nouns. So proper names seem to be unlike common nouns in at least this respect. Burge resolves the discrepancy by suggesting that bare, unmodified names actually have hidden determiners attached. A name such as *Fido*, when used on its own is, unbeknown to the speaker, actually the phrase *That Fido* or *The Fido* in disguise.

The rival view is Segal's contention that proper names are not common nouns but instead are a special kind of word, paired in each speaker's mind with a special kind of mental representation – an individual concept. These individual concepts are mental representations that encode information about the individuals named. So the name *David Bowie* is paired with an individual concept of David Bowie, perhaps containing the information that he sings, plays the saxophone, is married to a runway model, has probably had plastic surgery, and so on. Names, in Segal's account, are not at all like common nouns, encoding predicates that can apply to more than one person. Rather, they are labels that attach to conceptual information about particular individuals. There are not many David Bowies sharing one name. Rather, there are potentially many names *David Bowie*, each linked to a different individual concept.

Empirical Evidence

It might seem that in the end, the differences between the two theories do not amount to much. Burge says that the name *Fido* can be applied to anything that is a Fido, whereas Segal says that it only applies to one individual and that the reason why there seem to be so many Fidos is that there are many names for distinct individuals that happen to sound the same (call

these names *Fido*₁, *Fido*₂, etc.). Is there any real difference between these two theories? A behaviorist such as Quine might be inclined to think that, as long as each can be integrated into a larger theory of equal power in predicting linguistic behavior, then there is no fact of the matter about which is correct. But a cognitivist would rather suppose that there is a way to tell how the language system works, reflected in the biology and psychology of language, and that at most only one of the two suggestions can be correct. And it seems, at least at first glance, that the evidence from the task, algorithmic, and physical levels supports Segal's theory over Burge's.

At the task level, cognitivists consult the intuitions of native speakers to determine the characteristics of the language that they speak. In the case of proper names, the two theories under consideration make different predictions about the syntax of English. Burge's theory predicts that bare names actually have a hidden determiner word attached to them. But this view has some trouble accounting for common intuitions about how names and common nouns can be used. For example, why is it that determiners can go unpronounced when attached to names, but not when attached to common nouns, as shown by (13)? And why is it that sometimes special contexts are required to insert the determiner in front of a name? For example, to the question "Where do you live?" the response in (14a) seems natural whereas (14b) sounds awful. If *Saint Louis* is really short for a phrase such as *that Saint Louis*, then why can we not say (14b)?

(14a) I live in Saint Louis.

(14b) *I live in that Saint Louis.

At the algorithmic level, cognitivists look at psychological evidence regarding how linguistic knowledge is represented and processed. Again, the two theories make different predictions about the psychology of names. Burge predicts that names that sound the same are the same name, whereas Segal predicts that each individual's name is distinct. If Segal is right, there should be evidence that people tend to expect identical-sounding names to apply only to a single individual. Again, there is some evidence that supports Segal's prediction. It seems that children learning English as a first language expect there to be a class of nouns that refer to only one thing and make use of syntactic clues such as the presence or absence of determiners to decide whether to apply new words to other objects or not. For example, when told that a novel object is *wuzzle* (with no determiner), children are reluctant to apply the new word to other novel objects, even when they are highly similar to the original. But when told that the novel object is *a wuzzle*, they will happily generalize

the term to other objects that seem to share some salient properties with the original – just like ordinary common nouns.

Burge's theory also predicts that names are **not** a special kind of noun, whereas Segal predicts that they are. If Segal is right, we should expect to find psychological differences between names and common nouns. We might also expect some physical-level differences. (Recall that at the physical level, cognitivists look to neurological evidence for or against the kinds of representation and processing they propose in their algorithmic-level theories.) Again, the evidence seems to support Segal's view over Burge's. As previously noted, children seem to be prewired to look for names as well as common nouns. In addition, psychological studies on adults reveal that proper names are much harder to recall than common nouns, suggesting distinct storage and/or processing. And at the physical level, certain kinds of brain damage can cause people to lose their ability to use proper names while leaving their ability to use common nouns intact, and vice versa. This strongly suggests that names are a special kind of word stored in a separate area of the brain.

In fact, things are not as bad as all that for Burge's theory. Segal (2001), in his much more complete and sober account, correctly points out that the psychological and neurological evidence is still quite sketchy and open to interpretation. It is quite possible that a committed Burgian could find a way to keep the common noun theory of names alive. The main point of this example has been to show how, in principle, multidisciplinary evidence from all three levels of cognitive analysis can bear on an issue in semantics. Whereas a behaviorist might be content with two theories that are equally good at describing some aspect of linguistic behavior, the new philosopher of language looks deeper to try and find out which theory does a better job of accounting for the cognitive aspects of language.

Final Words

The work on proper names reviewed here nicely illustrates the main features of the new philosophy of language. Burge and Segal's truth-theoretic approach to semantics is as rigorously formal as any theory in the ideal language tradition; the attention to ordinary speaker intuitions in mediating between semantic theories echoes the approach of the ordinary language philosophers; the mentalistic nature of the theory, the formal, computational nature of truth theories, and the openness to evidence from all levels of cognitive analysis clearly places the work in the cognitivist tradition.

But is this new hybrid approach really philosophy of language, or is it just a branch of linguistics or psychology? There are still those who hold out the hope that analysis of language will eventually help with the resolution of issues in other branches of philosophy, even if only in providing a starting point, and most contemporary philosophers of language remain sensitive to the philosophical puzzles and paradoxes that drove the ideal and ordinary language philosophers. Indeed, one of the selling points of both Burge's and Segal's theories of proper names is that they can account for the meanings of empty names. But heeding Grice's lesson about the difficulties of determining what is said and heeding the lessons from contemporary linguistics about the complexities of ordinary language, few still believe any philosophical problem will be solved or dissolved with just a little bit of armchair reflection on conventions of use. The new philosophy of language promises progress on some of the difficult traditional problems in philosophy of language (and perhaps on more general philosophical problems) by combining careful conceptual analysis with detailed attention to empirical results from the scientific study of language, the mind, and the brain.

See also: Aristotle and Linguistics; Behaviorism: Varieties; Causal Theories of Reference and Meaning; Descriptions, Definite and Indefinite: Philosophical Aspects; Empty Names; Epistemology and Language; Externalism about Content; Implicature; Indeterminacy, Semantic; Innate Knowledge; Mentalese; Metaphysics, Substitution *Salva Veritate* and the Slingshot Argument; Objects, Properties, and Functions; Proper Names: Philosophical Aspects; Propositions; Reference: Philosophical Theories; Representation in Language and Mind; Semantics–Pragmatics Boundary; Tacit Knowledge; Thought and Language: Philosophical Aspects; Use Theories of Meaning.

Bibliography

- Adams F & Stecker R (1994). 'Vacuous singular terms.' *Mind and Language* 9(4), 387–401.
 Barber A (ed.) (2003). *The epistemology of language*. Oxford, UK: Oxford University Press.
 Burge T (1973). 'Reference and proper names.' *Journal of Philosophy* 70(14), 425–439.

- Chomsky N (1959). 'A review of B F Skinner's *Verbal Behavior*.' *Language* 35(1), 26–58.
 Chomsky N (1969). 'Quine's empirical assumptions.' In Davidson D & Hintikka J (eds.) *Words and objections: Essays on the work of W. V. Quine*. Dordrecht: D. Reidel. 53–68.
 Chomsky N (2000). *New horizons in the study of mind and language*. Cambridge, UK: Cambridge University Press.
 Dawson M R W (1998). *Understanding cognitive science*. Malden, MA: Blackwell.
 Fodor J A (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
 Fodor J A (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.
 Gazzaniga M S, Ivry R B & Mangun G R (2002). *Cognitive neuroscience: The biology of the mind, second edition*. New York: W. W. Norton.
 Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
 Jackendoff R (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
 Larson R K & Segal G (1995). *Knowledge of meaning: An introduction to semantic theory*. Cambridge, MA: MIT Press.
 Marr D (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.
 Quine W V O (1960). *Word and object*. Cambridge, MA: MIT Press.
 Russell B (1919). *Introduction to mathematical philosophy*. London: George Allen & Unwin.
 Russell B (1957). 'Mr. Strawson on referring.' *Mind* 66, 385–389.
 Ryle G (1949). *The concept of mind*. New York: Barnes & Noble.
 Segal G (2001). 'Two theories of proper names.' *Mind and Language* 16(5), 547–563.
 Skinner B F (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
 Soames S (2003). *Philosophical analysis in the 20th century* (vols 1–2). Princeton, NJ: Princeton University Press.
 Sperber D & Wilson D (1995). *Relevance: Communication and cognition*. Cambridge, MA: Blackwell.
 Strawson P F (1950). 'On referring.' *Mind* 59, 320–344.
 Trask R L (1999). *Language: The basics* (2nd edn.). New York: Routledge.
 Valentine T, Brennan T & Brédart S (1996). *The cognitive psychology of proper names: The importance of being earnest*. New York: Routledge.

Communication: Semiotic Approaches

E Rigotti and S Greco, University of Lugano, Lugano, Switzerland

© 2006 Elsevier Ltd. All rights reserved.

The Rise of a Controversy

The topic at issue in this article shows several problematic aspects. In communication research, we are presently crossing a phase of intensive innovation, in which the paradigm and the role of the different disciplines are changing remarkably. For a long time, the leading role in this area was played by the *sciences du langage*, in particular by semiotics and linguistics. Nowadays, this role is played by a complex epistemological interplay, where other human and social sciences – focusing on the organizational assets of communication context – as well as technological disciplines contribute to the study of real communicative events. Thanks to these contributions, it has become evident that real communicative events are not only influenced, but functionally governed by their actual context (enterprises, institutions, communities, and other social organizations . . .) and by the media, by which they are not only broadcasted, but also structured.

Moreover, even linguistic sciences, which are expected to explain the internal structure of a communicative event, are largely adopting a model of communication whose conceptual frame is no longer essentially semiotic, but rather pragmatic. The prevailing of a pragmatic paradigm seems to have strongly redimensioned the semiotic claim. More specifically, both major trends – Speech acts Theory and Relevance Theory (i.e., the ostensive inferential model of communication) – are proposing a vision of communication that does not focus on semiotic aspects.

While the former of these trends has developed its own model essentially ignoring the semiotic approach (Austin, 1962 and Searle, 1969), the latter has created a proper controversy, initiating a sort of campaign against the semiotic approach and its academic power.

At this point, it is useful to outline the ostensive-inferential model of Relevance theory synthetically, in order to specify its criticism of the semiotic model of communication, and also what it justly presupposes a semiotic model to be. In fact, the ostensive-inferential model, whose roots are Paul Grice's and David Lewis' works (Sperber and Wilson, 1995/1986: 2) is introduced by means of those aspects that oppose it to the semiotic model (Sperber and Wilson, 1986: 6) "The semiotic approach to communication (as Peirce called it and we will call it ourselves), or the semiological approach (as Saussure and his followers called it), is a

generalization of the code model of verbal communication to all forms of communication," and is thus to be abandoned, since it does not seem to explain the real functioning of communicative events: "The code model of verbal communication is only a hypothesis, with well-known merits and rather less well-known defects. [. . .] Its main defect, as we will shortly argue, is that it is descriptively inadequate: comprehension involves more than the decoding of a linguistic signal" (Sperber and Wilson, 1995/1986: 6). In other words, the semiotic approach appears to interpret communication as a process where a speaker constructs a message by coding a certain meaning by means of a linguistic system, and transfers it to a hearer who simply **decodes** it, thus retrieving its original meaning. The roles of the speaker and the hearer in a communicative event are thus reduced to coding and decoding respectively. The scholars of the ostensive-inferential approach to communication, relying on wide and unquestionable evidence, argue that the process of interpreting a message by the hearer is far more complex, and that the semiotic component represents a rather short stretch of the communicative process.

The semiotic component is neither necessary nor sufficient to explain the process of communication.

Firstly, it is not necessary because many messages do not make use of a linguistic system; very often, the communicator addresses the hearer not through words of a certain natural language, or through another semiotic system, but through traces by which the hearer is expected to be guided to infer the communicative intention of the message. Sperber and Wilson (1995/1986: 25) argued in fact that Grice's originality consisted in suggesting that the identification of the communicator's intentions is **sufficient** for the achievement of successful communication, and the mediation of a verbal code is not necessarily needed. The authors give an example (Sperber and Wilson, 1995/1986: 25–26) that shows how communication may succeed even without the help of the coding-decoding process. If Peter asks Mary: 'How are you feeling today?', Mary may answer by pulling a bottle of aspirin out of her bag and showing it to him. Although there is no code or convention that rules the interpretation of her behavior, this action can be taken as strong evidence that she wants to inform Peter that she does not feel well. In this sense, Mary and Peter can be said to have communicated, even if they have not made use of any verbal or nonverbal code.

The semiotic component is not even sufficient, even in the very usual cases where it is present. However, large it may be, interpretation requires that various

contextual aspects are involved in order to complete the information carried by the semiotic component: “Verbal communication is a complex form of communication. Linguistic coding and decoding is involved, but the linguistic meaning of an uttered sentence falls short of encoding what the speaker means: it merely helps the audience infer what she means” (Sperber and Wilson, 1995/1986: 27). Within this complex form of communication, the results of the decoding process are considered a piece of evidence from which the hearer, through a noncoded mechanism, can infer the speaker’s intentions. In this sense, the semiotic component becomes subservient to the inferential process. Using the terminology of Relevance Theory, an enrichment of the linguistic form of the message is however indispensable to obtain the semantic and pragmatic interpretation of a message. This is crucial to distinguish between ‘sentence’ and ‘utterance of a sentence.’ According to these authors, generative grammars fail to consider that a certain sentence may appear in an enormous variety of utterances that, though sharing a ‘core of meaning’ (Sperber and Wilson, 1995/1986: 9) bound to the linguistic code, each includes a different non-linguistic, context-bound meaning that can be neither predicted nor ‘calculated’ through a decoding process. Therefore, an inference process is required in order to grasp a complete representation of the communicator’s intentions. To give just an example (adapted from Sperber and Wilson, 1995/1986: 11), a sentence like ‘You’re leaving’ contains different levels of noncoded meaning: (1) an indexical (you) whose interpretation depends on the actual communicative event where the sentence is uttered; and (2) a set of possible interpretations: is the speaker informing the hearer that she is to leave? Is she making a guess? Or is she rather expressing disappointment because he is leaving?

Thus, the process of comprehension, through which the hearer reconstructs the communicator’s intentions, is not a decoding process, but rather an inferential process. Whereas the decoding process “starts from a signal and results in the recovery of a message which is associated to the signal by an underlying code” (Sperber and Wilson, 1995/1986: 13), an inferential process starts from a set of premises and reaches a conclusion warranted by the premises themselves. Among possible interpretations of an utterance, the hearer chooses the most adequate to certain expectations of truthfulness, informativeness, and comprehensibility. The inferential process of comprehension is an essential component of communication, which is nonetheless often integrated by the employment of a code. A common code between the interlocutors turns out to be the most powerful, however not indispensable, tool for communicating.

Sperber and Wilson’s critical remarks are generally convincing and acceptable, where they criticize the attempt to explain the interpretative process merely in terms of decoding. Less convincing is the more general criticism of all semiotic models of communication, accusing them of reducing communication to a coding and decoding process.

Saussurean ‘Signification’ as Keyword and Sign of Contradiction

Our thesis is that Sperber and Wilson’s criticism, which is legitimate in relation to certain semiotic models, is unacceptable for others. Furthermore, in our opinion, their reductive vision of the function of the semiotic component within a communicative process is by no means convincing.

For both points, we should briefly reconsider some of the communication models more or less explicitly proposed by semioticians and linguists in the past century. It is almost compulsory to start by referring to Ferdinand de Saussure, with whom the beginning of modern linguistics in its structuralist phase is usually connected. His representation of the communication process seems to constitute a typical coding and decoding model, **Figure 1**.

Here the speaker, having in mind a particular *signifié*, correlates it to the corresponding ‘signifiant’ of her linguistic system (*langue*), which is perceived by the hearer who correlates it to the correspondent *signifié* of the same linguistic system. Nonetheless, Saussure’s *Cours* has a problematic nature; its real function is to witness a deep and complex meditation rather than systematically expounding a theory. Thus, beyond the approximate presentation of the discourse circle (*circuit de la parole*), the Saussurean text introduces the fundamental but problematic distinction between *signifié* – defined as a meaning carried by an element of a *langue* (a linguistic system) – and *signification* – a term denoting a notion that remains rather opaque in the Saussurean text. Its interpretation and the evaluation of its role in Saussure’s doctrine is nevertheless crucial, and turned out to characterize the two main divergent trends that emerged within post-Saussurean structuralism. It is worth noticing that these two trends have developed



Figure 1 Ferdinand de Saussure’s model.

considerably different attitudes toward communication in their theoretical elaboration. Relevant representatives of both trends being numerous, only those scholars who cover significant and universally acknowledged cruxes will be mentioned here.

The Functionalist Reading

Let us consider Saussure's text. In a passage in chapter IV of the second part of the *Cours de linguistique générale* (1916/1995: 158–159), Saussure seems to employ the term *signification* as equivalent to *signifié*: *signification* seems to be nothing but the counterpart of the auditive image, “*un des aspects de la valeur linguistique*” or, better, the value of the conceptual component of the linguistic sign. Nonetheless, in the following passages, Saussure opposes *signification* to *signifié* throughout a series of interlinguistic confrontations (*mouton* vs. *sheep* and *mutton*, French plural vs. Sanskrit plural and dual, etc). So, without explicitly saying it, Saussure employs *signification* as opposed to *signifié*; interlinguistic comparisons between different language-bound *signifiés* are possible thanks to a conceptualization of reality that is formed somehow independently of these *signifiés*. This distinction lets us guess the existence of a complex correlation between the two semantic dimensions (reasonably understood as interpretation), which goes from the *signifiés* obtained through the coding to the *significations*, which articulate the *parole* (the speaker's actual message). Without this conceptualization, evoked by the use of the term *signification*, such a comparison between different languages would be simply impossible (on this point, see Rigotti & Rocci, in press). *Signification*, thus, has to be interpreted as an inter- or translinguistic category independent of the linguistic code, however correlated to it. If we integrate this notion into the Saussurean *circuit de la parole*, we obtain a more comprehensive model of communication, where the correlation of *signifiant* and *signifié* is only a stretch of a more complex path, starting with the actual meaning intended by the speaker, and ending with the reconstruction of this meaning tentatively operated by the hearer.

The interpretation of the Saussurean text presupposed by this model is explicitly adopted by N. Troubezko in his *Grundzüge der Phonologie* (1939), where the *signifiés* are considered, at the level of *langue*, as abstract rules and conceptual schemes, which need to be related to the actual *significations* emerging from language use (see also Rigotti & Rocci, in press: 5). On this point, M. Bréal (1844–1995: 552b) observes that, where we need to employ a certain word in communication, we ‘forget’ all possible meanings of that word except the one that corresponds to our thought (“*s'accorde avec notre pensée*”). Although the other meanings are

still somehow present to our mind, we choose the one that corresponds to the meaning we want to express – i.e. to the *signification*. Here, the relation between *signifiant* and *signifié* is certainly not a coding-and-decoding one, since it is mediated by the speaker's choice to her communicative intention. The same approach to communication may be found in Karl Bühler's *Organonmodell*, as outlined in his *Sprachtheorie* (1934). Among the numerous pages of Bühler's text, which could be useful to elucidate his position on this issue, one passage seems particularly revealing (1934: 63), where Bühler argues that no code can ensure the correct interpretation of the word ‘horse’ as it is used in a text, where it can refer to a single entity or to the species of horses in general. The use within a text is not “*morphologisch erkennbar*,” i.e., it cannot be decoded by means of morphological aspects of the language, neither in Latin, a language that does not foresee articles, nor in the Indo-Germanic, article-provided languages. What allows us to correctly interpret the use of the word ‘horse,’ is a ‘detective-attitude’ towards the context of the communicative event, which aims at evaluating what the speaker has in mind: “Man muss es detektivisch gleichsam dem Kontexte oder den Umständen der Sprechsituation entnehmen, ob der Sprecher das eine oder das andere im Auge hat und meint.” Moreover, an author to whom Bühler is quite indebted, Philipp Wegener, had also stressed the interpretative aspect of communication 50 years previously. Wegener argues that the hearer has the complex task of understanding the speaker's action; for this purpose, he has to figure out what the ‘goal’ of the communicative action may be. Comprehension of verbal messages is achieved through ‘inferences’ (*Schlüsse*), which rely both on the meaning of the verbal signs as well as the experience of reality. So, where experience is lacking, comprehension is impossible (Wegener, 1885–1991: 128). For instance, one could not understand a sentence such as ‘a whistle of the train, and my brother was gone,’ if one had no experience of a train setting off from a station.

If, speaking of a semiotic approach to communication, we refer to this research tradition in linguistics and semiotics, the criticism put forward by the scholars of Relevance Theory loses its bite: in fact, the process of communication is not referred to as a coding-decoding process within this trend. Rather, one should acknowledge that the process of communicative inference is constantly associated with the concept of interpretation.

Nor would the objection be acceptable that, in these models, inference only plays the subservient role of integrating the semiotic process. Here, it must be noticed that, if inference is acknowledged as

a necessary integration of the semiotic component, it follows that the semiotic component itself is not considered sufficient for the accomplishment of the communicative event. Therefore, the inferential component becomes essential for communication. More specifically, in this first tradition, neither the speaker's coding nor the hearer's decoding hold the supremacy in the communicative process; the crucial moment is rather when it becomes clear what the speaker intended to communicate, and the hearer understands it. As Bühler claims, on the backdrop of a Husserlian philosophical vision, language always appeals to the speaker's knowledge of reality; and each time we understand the meaning of a communicative event, we deeply and unavoidably rely on a 'reality-driven selection' (*sachgesteuerte Selektion*, Bühler, 1934: 65), which constitutes the core of communication. Not by chance, a large part of Bühler's research is devoted to the study of the specific semantic mechanism of the 'indexicals' or 'deictics' (*Zeigwörter*). This term refers to linguistic units and structures whose meaning is reconstructed through the identification of an aspect of the communicative situation (Bühler, 1934 see in particular p. 107).

Here, given the importance that Bühler attributes to reality in the process of communication, it becomes clear why he adopts a 'triadic' notion of the sign, which is rather innovative if we compare it to other

structuralist models. In his *Organonmodell* (1934: 24), the sign is conceived as an 'instrument' for communicating; and communication is interpreted pragmatically, as an action accomplished by the speaker and the hearer. According to Bühler (1934: 52), communication must be viewed as a human 'action,' vitally bound to other meaningful human behaviors. Communication is related to other actions, and is an action in itself. In particular, Bühler distinguishes between *Sprechhandlung* (1934: 53), which is the human activity of communicating, i.e., the Saussurean *parole*, as opposite to *Sprachgebilde* (the *langue*, 1934: 57); moreover, with the notion of *Sprechakt* (1934: 62), he focuses on a single communicative action, and with *Sprachwerk* (1934: 53), he denotes the linguistic products resulting from a single human action of communicating.

Within the model, the sign is related to the speaker (*Sender*), the addressee (*Empfänger*) and the objects and states of affairs in reality (*Gegenstände und Sachverhalte*). The sign is bound to each dimension by a specific relation: with regard to the speaker, the sign is a 'symptom,' bound by a relation of 'expression' (*Ausdruck*); with regard to the addressee, the sign is a 'signal,' and stands in the relation of appeal (*Appel*); and, finally, with regard to the object, the sign is a 'symbol,' and stands in the relation of 'representation.' The following diagram, **Figure 2**, illustrates Bühler's model (1934: 28):

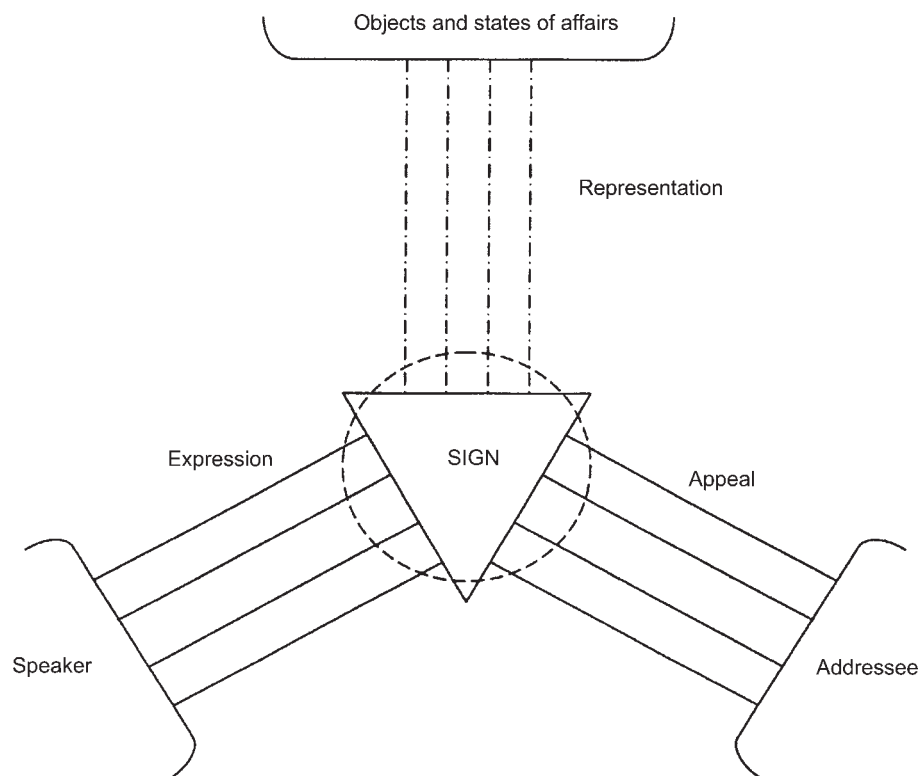


Figure 2 Karl Bühler's *Organonmodell*.

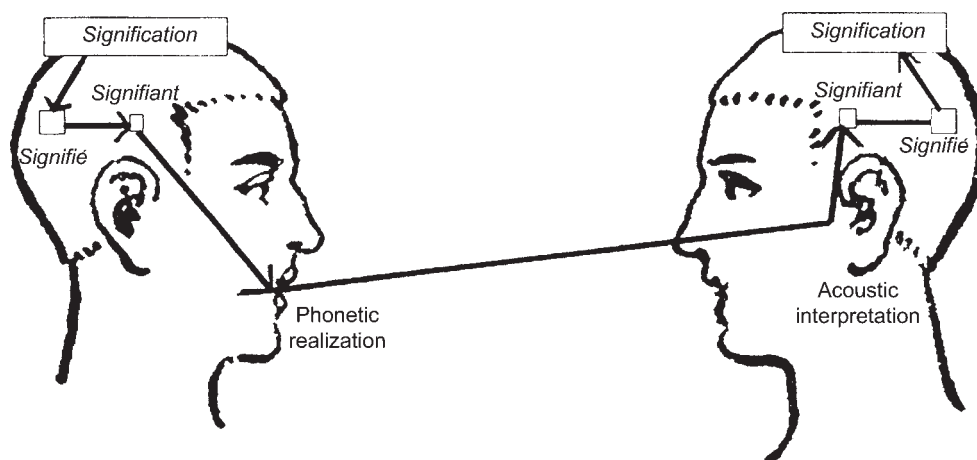


Figure 3 The model of communication within the functionalist reading.

The distinction between code dimension and discourse dimension of semantics, implied by the Saussurean terms *signifié* and *signification*, is tackled and deepened by another linguist: E. Benveniste, who introduced the terms ‘semiotic’ and ‘semantic’ (Benveniste, 1966a). He underlines that the content dimension of code units is a semiotic one; while the content dimension of the same units, insofar as they are used within a discourse, is truly semantic (on this point, see also Rocci, 2003). Moreover, among the indexicals investigated by Bühler, he focuses on personal pronouns, by which the communicative act and its constituents are mirrored in specific linguistic structures (Benveniste, 1966b). The study of personal pronouns on both the diachronic and the synchronic axes brings Benveniste to single out the essential role that is played by subjectivity (*I* and *You*) in communication.

On the basis of the Saussurean notion of ‘signification,’ conceived as the actual, situation-bound meaning of the sign in the communicative process, and of Bühler’s interpretation of the sign as an instrument for communicating, we could modify Saussure’s diagram and build a model of communication that is shared in its fundamental aspects by all the authors within the research tradition we have examined so far, Figure 3.

Even in its visual diversity, the well-known model proposed by Roman Jakobson (Jakobson, 1960/1995) is, in many respects, reminiscent of Bühler’s sign model. Being evidently influenced by Shannon and Weaver’s model, it brings to light the process of transmitting a message, thus offering a rather obvious metaphor of the communicative process, Figure 4.

Jakobson’s model has two indubitable merits: firstly, it takes into account, and represents synthetically, a complex set of factors; secondly, it deepens many of the specific functions of the message in relation to

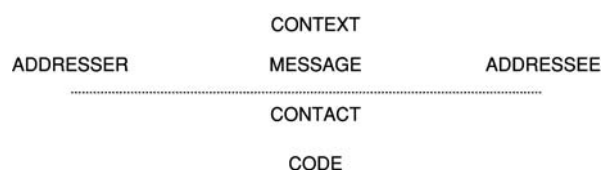


Figure 4 Roman Jakobson’s model of the fundamental factors of communication.

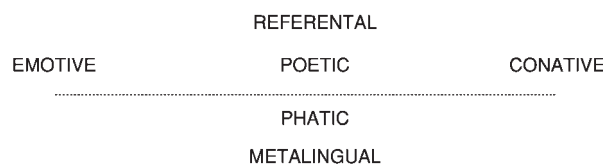


Figure 5 Roman Jakobson’s model of the textual functions.

each of these factors in the communication process. This Russian linguist treasures his former belonging to the significant experience of Russian formalism, by introducing the poetic function into his model, as an autotelic orientation of the message towards itself, Figure 5.

The graphic representation of Jakobson’s model appears to be richer than the sign scheme provided by Bühler. However, if we consider the model implicit in the theory of the latter, we have to recognize that Bühler’s model is richer in important respects: indeed, in Jakobson’s perspective, the pragmatic dimension is weakened; the essential role of inference in interpretation is ignored, as well as the relevance of context for interpretation. Another important aspect concerns the distinction between *signifié* and *signification*, reflecting the more general difference between language (*langue*) and speech (*parole*), which remains outside the graphic model outlined by Jakobson, even though it is adumbrated in some significant research (Jakobson, 1957).

Some Code-model Approaches

The precise definition of the Saussurean model represents a core issue for a large segment of semioticians of the past century. Indeed, besides the tradition we have tackled so far, another tradition of semiotic studies starts from a different interpretation of Saussure's *signification*. The second trend does not concentrate on the notion of *signification*, and therefore it does not focus on the textual and discursive dimension of the *parole*, whereas the point of view of the code (*langue*) is preferred. This position can be found not only in Hjelmslev's *Prolegomena to a theory of language* (1961), but also in various scholars belonging to French structuralism – among which R. Barthes plays a paradigmatic role – and in Umberto Eco's first semiotic theory, expounded in his work *Trattato di semiotica generale* (1979).

It is worth noticing that it is quite difficult to infer a model of communication from these positions. Barthes, for instance, stresses the interpretation of language as a system, whereby the individual performing a particular act of *parole* (a discourse) simply selects and actualizes one of the possible states of the system (Barthes, 1964). As the semantic dimension is exhaustively represented by the system of the *signifiés*, the meaning of communicative messages is not built by a speaker for an addressee, but it is rather one possible product the system can generate. The human subject is excluded from the communication process; communication itself, conceived as a communicative interaction between two human beings, i.e., as the junction of the communicative action of the speaker with the interpretative action of the addressee, fails to be considered at all. Umberto Eco (1979: 8) defined communication as “the passage of a signal (*not* necessarily a sign) from a source (through a transmitter, along a channel) to a destination.” This definition is meant to include both cases of machine-to-machine passages of information (see also 1979: 32), and cases where the destination (and not necessarily the source) is a human being. In the latter case, communication involves the process of signification, “provided that the signal is not merely a stimulus but arouses an interpretive response in the addressee” (1979: 32). The process of signification is not conceived as a communicative action; the focus here is on the signification system, “an autonomous semiotic construct that has an abstract mode of existence independent of any possible communicative act it makes possible” (1979: 9). Thus, it is the system that guarantees communication, and the existence of the system does not presuppose the existence of actual communicative events. On the contrary, communication between human beings necessarily presupposes a

signification system (thus excluding cases of nonverbal, ostensive communication). It must be observed that Eco explicitly discusses the problem of what the place of the human being, i.e., “the ‘acting subject’” (1979: 314) within semiotics should be. He concludes that what is outside the signification system – its “material expressions” (1979: 317) might even be “tremendously important,” but it is beyond the subject of semiotics. In fact, as Eco argues, the proper subject of signification is “nothing more than the continuously unaccomplished system of systems of signification that reflects back on itself,” whereas individual material subjects only “obey, enrich, change and criticize” the signification system (1979: 315).

As emerges from our survey of some theories within the second trend of Saussurean semiotics, speaking of a proper ‘model of communication’ in relation to them turns out to be quite difficult. In fact, communication in itself is intrinsically ignored. What they hypothesize are the mysterious workings of an autonomous semiotic program, which would auto-install and run on a mass of undifferentiated terminals, thus defining their individual or network sign production.

Charles Sanders Peirce

The model of communication of the first trend inspired by Saussurean semiotics, which we found in Bühler, and which is confirmed by recent pragmatic models, shows interesting analogies with another tradition, often considered as alternative to the Saussurean one: the semiotic model by Charles Sanders Peirce. As Bühler would do in the 1930s, Peirce had already proposed a triadic notion of sign at the end of the 19th century, **Figure 6**.

According to what Peirce wrote in 1897 (1897–1935–1958: 2.228), “A sign, or ‘representamen’ is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign

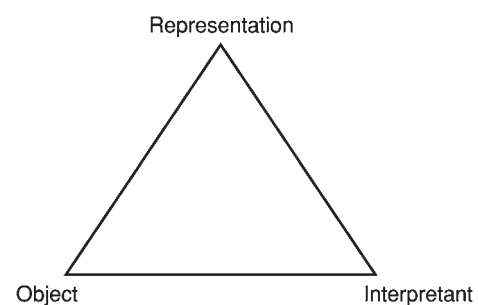


Figure 6 Charles Sanders Peirce's model of sign.

which it creates I call the interpretant of the first sign.” Although Peirce is often considered one of the founders of semiotics, it must not be forgotten that his contribution is particularly relevant from the logical and philosophical points of view. And his interest for semiotics concerns the cognitive rather than the communicative dimensions. Nonetheless, his contribution is also significant for semiotics and for a theory of communication. Concerning semiotics, we have to underline that Peirce’s notion of sign includes ‘symbols,’ as well as ‘indexes’ (bound to the object through a real connection), and ‘icons,’ which remind of objects by reproducing their features. Semiotics turns out to include both verbal and nonverbal dimensions. Nevertheless we should also consider that, within Peirce’s enormous scientific production, we find some significant cues for a significantly comprehensive communication model.

Firstly, the correlation of the sign with both subjectivities involved in communication is highlighted by the above-quoted definition, where the subject to whom the sign is addressed is explicitly mentioned and the addresser is presupposed. On this point, M. Hansen (2002) argued that Peirce’s approach implies an active involvement of the speaker and the addressee in the process of interpretation. In fact, the ‘representamen’ does not univocally imply a certain ‘interpretant,’ but it rather suggests several possible interpretations. Here, the interaction of the speaker and the addressee is necessary to evaluate the interpretation to be chosen: the context of interpretation is actively constructed by the interlocutors, on the basis of the experience of the knowledge community.

Secondly, we find in the Peircean text a truly pragmatic reading of the process of interpretation, as the ‘final interpretant’ of a sign is the ‘habit change,’ i.e., “a modification of a person tendencies toward action” (Peirce, 1897–1935–1958: 5.476; on this point, see also Rigotti & Rocci, 2001: 48).

Conclusive Remarks

We might conclude by arguing that the criticism moved against the semiotic tradition by the scholars of Relevance Theory is only valid for those semiotic approaches which can be defined as code-driven, and depend on a reductive interpretation of Saussure. They conceive of the sign as a binary unit, and thus reduce communication to a coding and decoding process. The criticism does not hold for all those, indeed rather numerous, approaches (Peirce and the functionalist interpretation of the Saussurean *Cours*: Troubetzkoy, Bühler, Jakobson, Bally, Sechehaye, Karcevskij ...), where a pragmatic (in the sense of the *Organonmodell* of language) and triadic representation of the sign

allows to understand communicative events in an adequately comprehensive perspective.

Our short survey of semiotic approaches to communication in the 20th century shows that not all of them can be considered as code-models, it also puts forward – this concerns in particular some authors like Peirce, Bühler, and Benveniste – the possibility and even the reasonableness of constructing a semio-pragmatic model of communication (Searle, 1969; Clark, 1996), concerning both the wording and the interpretation side, this latter being based on semiotic, metaphoric (Lakoff, 1980; Danesi, 2004) and inferential processing (Sperber and Wilson, 1995–1986); comprehending both verbal and nonverbal communication (Rocci, 2003), and including a theory of subjectivity as one of its relevant components (see Rigotti and Cigada, 2004). And the semiotic tradition of the 20th century could be shown to be helpful in this endeavor.

See also: Relevance Theory; Saussure: Theory of the Sign; Speech Acts.

Bibliography

- Austin J L (1962). *How to do things with words*. Oxford: Oxford University Press.
- Bally Ch (1950). *Linguistique générale et linguistique française*. Bern: Franke.
- Barthes R (1964). *Eléments de sémiologie*. Paris: Editions du Seuil.
- Benveniste E (1966a). ‘Les niveaux de l’analyse linguistique.’ In Benveniste E (ed.) *Problèmes de linguistique générale*. Paris: Gallimard.
- Benveniste E (1966b). ‘La subjectivité dans le langage.’ In Benveniste E (ed.) *Problèmes de linguistique générale*. Paris: Gallimard.
- Bréal M (1884–1995). ‘Comment les mots sont classés dans notre esprit.’ In Desmet P & Swiggers P (eds.) *De la grammaire comparée à la sémantique. Textes de Michel Bréal publiés entre 1864 et 1898*. Leuven/Paris: Peeters. 283–291.
- Bühler K (1934). *Sprachtheorie. Die Darstellungsfunktion der Sprache* (2nd edn.). Stuttgart/New York: Gustav Fischer.
- Clark H (1996). *Using language*. Cambridge: Cambridge University Press.
- Danesi M (2004). *Poetic logic: the role of metaphor in thought, language and culture*. Madison: Atwood Publishing.
- Eco U (1975). *Trattato di semiotica generale*. Milan: Bompiani. English translation (1976), *A theory of semiotics*. Bloomington: Indiana University Press.
- Grice H P (1957). ‘Meaning.’ *Philosophical Review* 66, 377–388.
- Hansen M B M (2002). ‘Sémiotique peircéenne et analyse des interactions verbales.’ In Andersen H L & Nölke H (eds.) *Macro-syntaxe et macro-sémantique. Actes du*

- colloque international d' rhus*, 17–19 mai 2001. Bern: Lang. 361–381.
- Hjelmslev L (1943–1961). *Prolegomena to a theory of language*. Madison: University of Winsconsin Press. [Omkring sprogteoriens grundlæggelse.]
- Jakobson R (1957). 'Shifters, Verbal Categories, and the Russian Verb.' In Rudy S (ed.) *Selected Writings*. The Hague: Mouton.
- Jakobson R (1960–1995). 'The speech event and the function of language.' In Waugh L R & Monville-Burston M (eds.) *On language*. Cambridge/London: Harvard University Press.
- Karcevskij S O (1929). 'Du dualisme asymmetrique du signe linguistique.' *Travaux du cercle linguistique de Prague* 1, 88–93.
- Lakoff G (1980). *Metaphors we live by*. Chicago: The University of Chicago Press.
- Lewis D (1969). *Convention*. Cambridge: Harvard University Press.
- Peirce Ch S (1897–1935–1958). *Collected papers of Charles Sanders Peirce* (8 vols). Hartshorne C & Weiss P (eds.). Cambridge: The Belknap Press of Harvard University Press.
- Rigotti E & Cigada S (2004). *La comunicazione verbale*. Milano: Apogeo.
- Rigotti E & Rocci A (2001). 'Sens – non-sens – contresens.' *Studies in communication sciences* 1, 45–80.
- Rigotti E & Rocci A (in press). 'Le signe linguistique comme structure intermédiaire.' In Saussure L de (ed.) (in press). *Nouvelles perspectives sur Saussure. Mélanges offerts à René Amacker, Publications du Cercle Ferdinand de Saussure*. Genève: Droz.
- Rocci A (2003). 'La testualità.' In Bettetini G, Cigada S, Raynaud S & Rigotti E (eds.) *Semiotica II. Configurazione disciplinare e questioni contemporanee*. Brescia: La Scuola. 257–319.
- Secheaye A (1926). *Essay sur la structure logique de la phrase*. Paris: Champion.
- Saussure F de (1916/1995). 'Cours de linguistique générale.' In Bally Ch & Sechehaye A (ed.) with the collaboration of Riedlinger A. Paris: Payot.
- Searle J (1969). *Speech acts*. Cambridge: Cambridge University Press.
- Sperber D & Wilson D (1995/1986). *Relevance: communication and cognition* (2nd edn.). Blackwell Publishers.
- Troubetzkoy N S (1939). *Grundzüge der Phonologie*. Prague: Travaux du Cercle Linguistique de Prague.
- Wegener Ph (1885–1991). 'Untersuchungen über die Grundfragen des Sprachlebens.' In Knobloch C & Koerner K (eds.). Amsterdam/Philadelphia: John Benjamins.

Communication, Understanding, and Interpretation: Philosophical Aspects

D Hunter, State University of New York at Buffalo, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Philosophers have asked questions about both the nature and the extent of communication. What, for instance, is the difference between expressing a belief and communicating one, and what role does language play in communication? What must one know to interpret another person, and can we ever really understand another person? Philosophers are still at work refining these questions and considering answers. This discussion aims to sketch the direction of this work.

The Nature of Communication

There might seem nothing especially puzzling about the nature of communication. After all, we seem to communicate with each other all the time and succeed, largely without having to reflect on the nature of our success. We tell others what we think or believe, and they in turn tell us what they think or believe. One way to begin to see what philosophers

have found puzzling about the nature of communication is to consider whether nonhuman animals – dogs and cats, for instance – ever communicate with each other or with us.

Suppose that Fido knocks his food bowl over whenever he is hungry, and he does it only in the presence of his master, whom he then looks at intently. Is Fido trying to communicate with his master? To answer this, we need to know more about what it is to try to communicate. One initially helpful distinction is between communicating a belief and merely manifesting or revealing one. Agents manifest or reveal something about their beliefs, desires, and intentions whenever they act, and ordinarily we can explain and predict these actions by reference to what (we think) they believe, want, and intend. This applies as much to dogs and cats as to people. We might, for instance, speculate that Fido is knocking his bowl over in part because he knows it is empty and wants it to be full. We might even, somewhat more ambitiously perhaps, say that Fido knows that if he knocks the bowl around his master will fill it for him. These attitudes are manifested by Fido's actions, in the sense that we can infer from the actions that he has these beliefs and desires. But, of course, it is one

thing for an action to reveal what one believes or desires and another to communicate what one believes or desires. After all, sometimes our actions reveal more than we want, as when we let something slip or when someone is spying on us. Genuine communication, it seems, requires **intending** to reveal what one believes; it requires manifesting one's beliefs or desires on purpose. Communication is thus an intentional activity. Did Fido intend to communicate his beliefs and desires when he knocked the bowl over?

Part of the difficulty in answering this stems from an ambiguity in the phrase 'communicate what he believes'. In one sense, for Fido to communicate what he believes is for him to communicate, in this case, that his bowl is empty, that it should be full, and that he is hungry. In this sense, what is communicated are the facts, or at least the facts as Fido takes them to be. But in another sense, for Fido to communicate what he believes is for him to communicate the fact that he has those beliefs and desires. In this sense, what is communicated is the fact that Fido takes or wants the world to be some way. Plausibly, to communicate what one believes or desires in the second sense, one must be aware that one has those beliefs and desires; one might doubt whether dogs and cats are aware of their own beliefs and desires. To doubt this is not to doubt whether dogs and cats have beliefs and desires or whether their actions are caused by their beliefs and desires. One can admit all of this while still doubting whether dogs and cats have the level of self-awareness necessary for intending to reveal that they have certain beliefs and desires. It is an empirical question whether dogs and cats have this level of self-awareness. In any event, the two senses of 'communicate what one believes' differ only over what is communicated and not over what it is to communicate. So even if it is true that Fido is incapable of communicating **that he believes that his bowl of food is empty**, it might still be true that he can communicate **that his bowl of food is empty**.

But before we can decide this, there is still more we need to know about what Fido is trying to do. More specifically, is Fido trying through his action to influence his master's beliefs or actions? Is Fido trying to get his master to believe something, or is he trying to get him to do something? To try to influence a person's beliefs requires, it seems, some awareness that that person has beliefs that can be influenced. I think it is clear that this is something that humans do when they communicate with one another. But if one doubted that dogs and cats are aware of their own beliefs and desires, then one would likely doubt that they are aware that people have beliefs and desires. Still, we sometimes communicate with others with the intention of influencing their actions, as when we give

warnings or orders. But perhaps we do this with the primary intention of influencing their beliefs, hoping that this will lead to the desired action. What is clear, in any event, is that we humans often do communicate with the intent of influencing the beliefs and desires of other people, whether or not this intention is required for genuine communication.

It is also clear that, whether it is necessary or not, acting with an intention to influence someone's beliefs is not enough for genuine communication. I might leave the milk carton on the counter, intending that when my wife notices that it is empty she will plan to buy one on her way home. If my plan succeeds, I will have influenced her beliefs. But it seems wrong to say that I would have communicated to her that the milk carton is empty, or that we need more milk, even though I deliberately caused her to believe these things. If, instead, I had made a show of holding the carton upside down in her presence I might well have communicated these things to her. But what is the relevant difference? Part of the difference, in this example anyway, is that in the second case my wife would know that I am trying to influence her beliefs. She would recognize that in making a show of holding the carton upside down I was trying to make her see that the milk is gone (or that I believe that it is gone). Genuine communication, it seems, may require that the audience recognize one's intention to influence his or her beliefs or actions.

This gives rise to a possible asymmetry in the case of dog-human communication. Plausibly, in knocking his food bowl around, Fido is intending to show his master that the bowl is empty, and no doubt his master can recognize this intention. This could be so even if Fido himself is not aware of having this intention or of his master's recognizing it. And of course the master can act with the intention of influencing Fido's beliefs or actions. But can Fido, or any dog, recognize such an intention in his master? Can Fido figure out that when his master pulls back on the leash he or she is trying to get Fido to heel? Again, if one doubts that dogs and cats are aware of their own beliefs and intentions then one might well doubt whether they are aware of the beliefs and intentions of others. And if they cannot recognize intentions and beliefs in others, and if this recognition is needed for the audience in genuine communication, then dogs and cats cannot be the audience of genuine communication. But so long as trying to communicate does not itself require being aware of intentions, then dogs and cats might be able to communicate, even if they are incapable of being the audience of a communication. This is the potential asymmetry.

One might think – perhaps with some justification – that the question of whether Fido is communicating

or not is, at this point in the discussion, more than a little terminological. After all, all sides can agree – supposing certain empirical questions about the self-awareness of dogs and cats to have been settled – about what dogs and cats and humans can and cannot do to try to influence beliefs and actions. Deciding whether to call what nonhuman animals do ‘communication’ may seem less important than recognizing the differences and similarities between what all sides agree that human and nonhuman animals can do. In any event, progress in understanding animal communication requires further empirical study, not terminological decision.

The discussion until now has left language out of the picture. The examples have all been of nonlinguistic communication. It is undeniable that we do communicate nonlinguistically with others using waves, winks, and kicks under the table (although there are terminological questions about just how to draw the line between linguistic and nonlinguistic communication). When we do so, we hope that our audience will be able to recognize our intention to communicate. There is nothing, I think, essentially new about linguistic communication except that it involves speech acts – acts done with words having a conventional meaning. However, reliance on conventions is not unique to linguistic communication, since nonlinguistic communication using signals and codes may also involve conventions. There is considerable current research about the nature or essence of human language and how it differs from codes or signal systems. Some leading philosophers also question whether human language is in any interesting sense meant for or designed for communication.

Some philosophers have argued, though, that nonlinguistic forms of communication are in a way dependent on linguistic forms. Some, such as René Descartes in the 17th century and Donald Davidson in the 20th century, held that genuine communication is essentially linguistic. With this view, having beliefs and intentions requires having language, and since (as we have seen) communication requires having beliefs and intentions, communication requires having language. Since dogs and cats have no beliefs they are, on this view, incapable of communication. This is, however, a minority opinion.

The rough definition of communication I have sketched applies just as much to linguistic communication as to nonlinguistic communication. To communicate is to perform an action, perhaps a speech act, with the intention of influencing an audience’s beliefs or actions and whose success requires that the audience recognize this intention. This is just a rough sketch of a complete picture. Considerable ongoing philosophical research is aimed at filling in

the details. In particular, research is focused on the precise nature of the relevant intentions.

Interpretation and Understanding

Communication, whatever it is, succeeds only when the audience correctly understands the communicator, when he or she correctly interprets what the communicator intended him or her to come to believe or do. Philosophers have asked various epistemological questions about the extent to which we can and do understand each other. Some are quite skeptical that communication ever succeeds.

One kind of philosophical question concerns the possibility that some of an agent’s thoughts are essentially private, in the sense that only that agent can think them. Such thoughts would be essentially incommunicable, ones an agent could never communicate to anyone else. One purported kind of example includes thoughts about the character of an agent’s own conscious experience. If no one else can know what it is like for me to taste chocolate or to see red, perhaps no one else can truly understand what I say when I try to describe these experiences. Perhaps what it is like to be me, from the inside, is something I can never fully communicate with another. A related kind of purportedly private thoughts are so-called first-person thoughts: thoughts an agent has about her own place in the world. Perhaps what I think when I think that I am in Buffalo is not what you (or I) think when you (or I) think that David Hunter is in Buffalo. Perhaps thoughts that locate my own position for me are not thoughts that others can share. If some thoughts are private in this way, then they would mark one principled limit to communication. But what such private thoughts might be like – and indeed whether they are even possible – are areas of ongoing philosophical research.

A more generalized skepticism about communication derives from the fact that what a person means by his or her words can never neatly be separated from what he or she believes. We use our words to express our beliefs, but what we intend to say depends on what we believe our words can be used to say. So we cannot understand what someone is saying without knowing what they believe, but our best insight into what they believe is through our understanding of what they say. This fact about the interdependence of meaning and belief has led some philosophers to suggest that what a person means depends on their entire cultural milieu. Different cultures, according to this position, have different systems of belief, or different worldviews, and interpreting or understanding an agent from another culture requires sharing or at least knowing that worldview.

A related position is that scientists working within different scientific paradigms, such as pre- and post-Einsteinian physics, cannot genuinely understand each other, because the meanings of their shared words derive from different theoretical structures. It is not just that what Newton meant by 'energy' is not what Einstein meant by it. Rather, the claim is, Einstein could not even understand what Newton meant by it, since he did not share Newton's scientific paradigm.

This skepticism conflicts with the apparent ease of cross-cultural communication and ordinary interpretation. Perhaps this skepticism rests on mistaken semantic assumptions. But it may be that the appearance of easy communication stems from the fact that we typically assume that other people generally share our beliefs and meanings. Perhaps this 'principle of charity' in interpretation creates an illusion of successful communication. In any event, the general point that there is some interdependence between what a person means by their words and what they believe can hardly be doubted and is enough to raise some doubt about just how successful ordinary communication really is.

A more severe skeptical worry starts from the fact that any theory is under-determined by evidence. It is a general fact about the nature of theories that very different, even conflicting, accounts of some phenomena will be compatible with all the available evidence. In the case of communication, this means that very different interpretations of someone's speech act will make equally good sense of all available (indeed, of all possible) evidence. Just as a scientific theory can be adjusted in countless ways to accommodate new evidence, so our interpretation of a speech act can be varied in countless ways by varying our interpretation of the agent's beliefs or meanings. By itself, this under-determination suggests that we might never be in a position to know that our interpretations are correct, since no amount of evidence could identify a single best interpretation. Meaning might forever transcend our ability to know. Some doubt whether this brand of skepticism constitutes a special problem for communication, since all of our theories are under-determined by evidence. Perhaps if our epistemic position with respect to meaning is no worse than that with respect to, say, atomic physics, then we can live with this much skepticism about interpretation.

However, the American philosopher W. V. O. Quine argued powerfully that there is a special problem in the case of communication. In early work, he stressed the idea that the under-determination of translation would occur, even if there were no under-determination of physics. Even if we agreed on all the physical facts, we might not, he argued, be able to agree on a

unique best interpretation of an agent's speech act, since there would still be room to vary the agent's beliefs and desires. The physical facts, in his view, do not determine the semantic ones. In later work, Quine stressed the idea that in the case of, say, physics, we are prepared to admit that the physical facts might transcend our cognitive capacities. We might, he held, simply lack the intellectual resources needed to discover those facts, since nothing in the facts themselves guarantees that we can know them. As a result, he said, we are prepared to say that even though conflicting physical theories might be equally compatible with all available evidence, at most one of them can be true. This means that we might not be able to tell which theory is true. However, Quine argued, it makes no sense to suppose that what someone means by his or her words or what he or she believes could transcend the evidence we have at our disposal. Facts about meaning and belief are, he held, essentially public and knowable by us. So, he concluded, the special problem for interpretation and understanding is that it makes no sense to say that one interpretation is truer than any other, so long as they each make equally good sense of the evidence. It is not that facts about meaning could go beyond what we can know; it is that there is nothing more to meaning than what we can know. And because what we can know fails to determine a unique translation, this means that translation is indeterminate.

While Quine's writings on this topic have been extremely influential, there is little consensus about just what his arguments are, let alone what the consequences of his view would be. And while he has won few converts, there is no consensus about where his arguments go wrong. Some have responded that Quine unjustly adopted behaviorist limits on the available evidence or that he overlooked other sources of evidence at our disposal. Still, the thesis of the indeterminacy of translation is one of the most significant contributions to the philosophy of communication of the 20th century.

See also: Behaviorism: Varieties; Causal Theories of Reference and Meaning; Conventions in Language; Empiricism; Epistemology and Language; Indeterminacy, Semantic; Radical Interpretation, Translation and Interpretationalism.

Bibliography

- Bach K & Harnish R M (1979). *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.
 Castaneda H (1966). "'He": a study in the logic of self-consciousness.' *Ratio* 8, 130-157.

- Chomsky N (1969). 'Quine's empirical assumptions.' In Davidson D & Hintikka J (eds.) *Words and objections*. Dordrecht: Reidel.
- Davidson D (1984). *Inquiries into truth and interpretation*. Oxford: Oxford University Press.
- Davidson D (1984). 'Thought and talk.' In Davidson D (ed.) *Inquiries into truth and interpretation*. Oxford: Oxford University Press.
- Dummett M (1993). 'Language and communication.' In *The seas of language*. Oxford: Oxford University Press.
- Frege G (1918/1977). 'Thoughts.' Geach P T & Stoothoff R H (trans.). In Geach P T (ed.) *Logical investigations*. New Haven: Yale University Press. xxx-xxx.
- George A (1986). 'Whence and whither the debate between Quine and Chomsky?' *Journal of Philosophy*.
- Grice H P (1967/1989). 'Logic and conversation.' In Grice H P (ed.) *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Kripke S (1982). *Wittgenstein on rules and private language*. Cambridge, MA: Blackwell.
- Kuhn T (1962). *The structure of scientific revolutions* (2nd edn., enlarged). Chicago: University of Chicago Press.
- Perry J (1993). *The problem of the essential indexical*. Oxford: Oxford University Press.
- Quine W V O (1960). *Word and object*. Cambridge, MA: MIT Press.
- Quine W V O (1992). *Pursuit of truth*. Cambridge, MA: Harvard University Press.
- Searle J (1969). *Speech acts*. Cambridge: Cambridge University Press.
- Sperber D & Wilson D (1988). *Relevance: communication and cognition*. Cambridge, MA: Harvard University Press.
- Whorf B J (1956). *Language, thought and reality*. Carroll J M (ed.). Cambridge, MA: MIT Press.
- Wittgenstein L (1953). *Philosophical investigations*. Anscombe G E M (trans.). Oxford: Blackwell.

Comparatives: Semantics

C Kennedy, Northwestern University, Evanston, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The ability to establish orderings among objects and make comparisons between them according to the amount or degree to which they possess some property is a basic component of human cognition. Natural languages reflect this fact: all languages have syntactic categories that express gradable concepts, and all languages have designated comparative constructions, which are used to express explicit orderings between two objects with respect to the degree or amount to which they possess some property (Sapir, 1944).

In many languages, comparatives are based on specialized morphology and syntax. English exemplifies this type of system. It uses the morphemes *more-er*, *less*, and *as* specifically for the purpose of establishing orderings of superiority, inferiority, and equality, respectively, and the morphemes *than* and *as* to mark the standard against which an object is compared:

- (1a) Mercury is closer to the sun than Venus.
- (1b) The Mars Pathfinder mission was less expensive than previous missions to Mars.
- (1c) Uranus doesn't have as many rings as Saturn.

In the case of properties for which specific measure units are defined, it is also possible to express differences between objects with respect to the degree to which they possess some property, even when the

predicate from which the comparative is formed does not permit explicit measurement:

- (2a) Mercury is 0.26 AU closer to the sun than Venus.
- (2b) ??Mercury is 0.46 AU close to the sun.

Languages such as English also allow for the possibility of expressing more complex comparisons by permitting a range of phrase types after *than* and *as*. For example, (3a) expresses a comparison between the degrees to which the same object possesses different properties, (3b) compares the degrees to which different objects possess different properties, and (3c) relates the actual degree that an object possesses a property to an expected degree.

- (3a) More meteorites vaporize in the atmosphere than fall to the ground.
- (3b) The crater was deeper than a 50-story building is tall.
- (3c) The flight to Jupiter did not take as long as we expected.

Finally, many languages also have related degree constructions that do not directly compare two objects but instead provide information about the degree to which an object possesses a gradable property by relating this degree to a standard based on some other property or relation. The English examples in (4) using the morphemes *too*, *enough* and *so* exemplify this sort of construction.

- (4a) The equipment is too old to be of much use to us.
- (4b) Current spacecraft are not fast enough to reach the speed of light.

- (4c) The black hole at the center of the galaxy is so dense that nothing can escape the pull of its gravity, not even light.

Example (4b), for example, denies that the speed of current spacecraft is as great as the speed required to equal the speed of light.

Gradability

A discussion of the semantics of comparison must begin with the semantics of gradable predicates more generally. Not all properties can be used in comparatives, as shown by the contrast between the examples in (1) and (5).

- (5a) ??Giordano Bruno is deader than Galileo.
 (5b) ??The new spacecraft is more octagonal than the old one.
 (5c) ??Carter is as former a president as Ford.

The crucial difference between predicates such as *expensive* and *close*, on the one hand, and *dead*, *octagonal*, and *former*, on the other, is that the first, but not the second, are gradable – they express properties that support (nontrivial) orderings. Comparatives thus provide a test for determining whether a predicate is inherently gradable or not.

The most common analysis of gradable predicates assigns them a unique semantic type that directly represents their order-inducing feature; they are analyzed as expressions that map their arguments onto abstract representations of measurement, or scales. Scales have three crucial parameters, the values of which must be specified in the lexical entry of particular gradable predicates: a set of degrees, which represent measurement values; a dimension, which indicates the property being measured (cost, temperature, speed, volume, height, etc.); and an ordering relation on the set of degrees, which distinguishes between predicates that describe increasing properties (e.g., tall) and those that describe decreasing properties (e.g., short) (see Sapir, 1944; Bartsch and Vennemann, 1973; Cresswell, 1977; Seuren, 1978; von Stechow, 1984a; Bierwisch, 1989; Klein, 1991; Kennedy, 1999; Schwarzschild and Wilkinson, 2002).

The standard implementation of this general view claims that gradable predicates have (at least) two arguments: an individual and a degree. Gradable predicates further contain as part of their meanings a measure function and a partial ordering relation such that the value of the measure function applied to the individual argument returns a degree on the relevant scale that is at least as great as the value of the degree argument. The adjective *expensive*, for example, expresses a relation between an object x and a degree of cost d such that the cost of x is at least as great as d .

In order to derive a property of individuals, it is necessary to first saturate the degree argument. In the case of

the positive (unmarked) form, the value of the degree argument is contextually fixed to an implicit norm or standard of comparison, whose value may vary depending on a number of different contextual factors (such as properties of the subject, the type of predicate, and so forth). For example, the truth conditions of a sentence such as (6a) can be represented as in (6b), where *size* is a function from objects to degrees of size and d_s is the contextually determined standard – the cutoff point for what counts as large in the context of utterance.

- (6a) Titan is large.
 (6b) $size(t) \geq d_s$

In the context here, the various objects in the solar system, the value of d_s is typically such that (6a) is false. If we are talking about Saturn's moons, however, then d_s is such that (6a) is true. This sort of variability is a defining feature of gradable adjectives as members of the larger class of vague predicates.

Comparison

In contrast to the positive form, comparatives (and degree constructions in general) explicitly fix the value of the degree argument of the predicate. There are a number of implementations of this basic idea (see von Stechow, 1984a, for a comprehensive survey), but most share the core assumption that the comparative morphemes fix the value of the degree argument of the comparative-marked predicate by requiring it to stand in a particular relation – $>$ for *more*, $<$ for *less*, and \geq for *as* – to a second degree, the comparative standard, which is provided by the comparative clause (the complement of *than* or *as*).

One common strategy is to assign the comparative morpheme essentially the same semantic type as a quantificational determiner – it denotes a relation between two sets of degrees. One of these sets is derived by abstracting over the degree argument of the comparative predicate; the second is derived by abstracting over the degree argument of a corresponding predicate in the comparative clause. This analysis presupposes that the comparative clause contains such a predicate. In some cases, it is present in the surface form (see (3b)), but typically, in particular whenever it is identical to the comparative predicate, it is eliminated from the surface form by an obligatory deletion operation.

For example, in the analysis developed in Heim (2000), *more (than)* denotes a relation between two sets of degrees such that the maximal element of the first (provided by the main clause) is ordered above the maximal element of the second (provided by the comparative clause). At the relevant level of semantic representation, a sentence such as (7) has the constituency indicated in (8a) (where material elided

from the surface form is struck through) and the truth conditions in (8b).

- (7) Titan is larger than Hyperion.
 (8a) [Titan is d large] more than [Hyperion]
 (8b) $\max\{d \mid \text{large}(t) \geq d\} > \max\{d' \mid \text{large}(h) \geq d'\}$

Note that because the truth conditions of the comparative form do not involve reference to a contextual norm, the comparative does not entail the corresponding positive. Thus (8a), for example, can be true even in a context in which (6a) is false.

Differential comparatives such as (2a) can be accounted for by modifying the basic semantics to include a measure of the difference between the respective (maximal) degrees contributed by the two arguments of the comparative morpheme (von Stechow, 1984a; Schwarzschild and Wilkinson, 2002). Such differences always correspond to closed intervals on a scale and so are measurable even if the degrees introduced by the base-gradable predicate themselves are not (Seuren, 1978; von Stechow, 1984b; Kennedy, 2001).

Because the standard of comparison is derived by abstracting over a degree variable in the comparative clause, this approach allows for the expression of arbitrarily complex comparisons such as those in (3). There are some limits, however. First, the comparative clause is a *wh*-construction, so the syntactic operation that builds the abstraction structure is constrained by the principles governing long-distance dependencies (see Kennedy, 2002, for an overview). Second, it is also constrained by its semantics; because the comparative clause is the argument of a maximalization operator, it must introduce a set of degrees that has a maximal element. Among other things, this correctly predicts that negation (and other decreasing operators) are excluded from the comparative clause (von Stechow, 1984a; Rullmann, 1995):

- (9a) ??Venus is brighter than Mars isn't.
 (9b) $\max\{d \mid \text{bright}(v) \geq d\} > \max\{d' \mid \neg \text{bright}(m) \geq d'\}$

The set of degrees d' such that Mars is not as bright as d' includes all the degrees of brightness greater than the one that represents Mars's brightness. Because this set has no maximal element, the maximality operator in (9b) fails to return a value.

The hypothesis that the comparative clause is subject to a maximalization operation has an additional logical consequence (von Stechow, 1984a; Klein, 1991; Rullmann, 1995); for any (ordered) set of degrees D and D' , if $D \subseteq D'$, then $\max(D') \geq \max(D)$. The comparative clause is thus a downward-entailing context and so is correctly predicted to license negative-polarity items and conjunctive interpretations of

negation (Seuren, 1973; Hoeksema, 1984; but cf. Schwarzschild and Wilkinson, 2002):

- (10a) The ozone layer is thinner today than it has ever been before.
 (10b) We observed more sunspot activity in the last 10 days than anyone has observed in years.
 (11a) Jupiter is larger than Saturn or Uranus. \Rightarrow
 (11b) Jupiter is larger than Saturn, and Jupiter is larger than Uranus.

Finally, the assumption that the comparative is a type of quantificational expression leads to the expectation that it should participate in scopal interactions with other logical operators. The ambiguity of (12), which has the (sensible) *de re* interpretation in (13a) and an (unlikely) *de dicto* interpretation in (13b), bears out this prediction.

- (12) Kim thinks Earth is larger than it is.
 (13a) $\max\{d \mid \text{think}(\text{large}(e) \geq d) (k)\} > \max\{d' \mid \text{large}(e) > d'\}$
 (13b) $\text{think}(\max\{d \mid \text{large}(e) \geq d\}) > \max\{d' \mid \text{large}(e) > d'\} (k)$

The extent to which comparatives interact with other operators and the implications of such interactions for the compositional semantics of comparatives and gradable predicates is a focus of current investigation (see Larson, 1988; Kennedy, 1999; Heim, 2000; Bhatt and Pancheva, 2004).

Comparison Cross-Linguistically

As previously noted, there are in fact several distinct semantic analyses of comparatives that differ in their details but share the core assumption that gradable adjectives map objects to ordered sets of degrees. For example, one alternative analyzes the truth conditions of a sentence such (7) as in (14); roughly, there is a degree d such that Titan is at least as large as d but Hyperion is not as large as d (Seuren, 1973; Klein, 1980; Larson, 1988).

- (14) $\exists d[(\text{large}(t) \geq d) \wedge \neg(\text{large}(h) \geq d)]$

Analysis (14) does not express an explicit ordering between two degrees but instead takes advantage of the implicit ordering on the scale of the predicate to derive truth conditions equivalent to (8b) – given the inherent ordering, (14) holds whenever the maximal degree of Titan's largeness exceeds that of Hyperion (and vice versa).

The fact that the underlying semantics of gradable predicates supports multiple equivalent logical analyses of comparatives appears at first to be a frustrating obstacle to the discovery of the 'right' semantics of the comparative. In fact, however, this

may be a positive result when we take into account the extremely varied syntactic modes of expressing comparison in the world's languages (see Stassen, 1985), which include forms that superficially resemble the logical representation in (14), such as the example from Hixkaryána in (15).

- (15) Kaw-ohra naha Waraka, kaw naha Kaywerye
tall-NOT he-is Waraka tall he-is Kaywerye
'Kaywerye is taller than Waraka'

Although it may turn out to be difficult to find clear empirical evidence to choose between competing, equivalent logical representations of comparatives within a particular language such as English, it may also turn out that a study of the various expressions of comparison in different languages will show that all the possible options provided by the underlying semantics of gradability are in fact attested. Comparatives, therefore, provide a potentially fruitful and important empirical domain for investigating broader typological questions about the mapping between (universal) semantic categories and (language-specific) syntactic ones.

See also: Monotonicity and Generalized Quantifiers; Negation: Semantic Aspects; Quantifiers: Semantics.

Bibliography

- Bartsch R & Vennemann T (1973). *Semantic structures: A study in the relation between syntax and semantics*. Frankfurt: Athänum Verlag.
- Bhatt R & Pancheva R (2004). 'Late merger of degree clauses.' *Linguistic Inquiry* 35, 1–46.
- Bierwisch M (1989). 'The semantics of gradation.' In Bierwisch M & Lang E (eds.) *Dimensional adjectives*. Berlin: Springer-Verlag. 71–261.
- Cresswell M J (1977). 'The semantics of degree.' In Partee B (ed.) *Montague grammar*. New York: Academic Press. 261–292.
- Heim I (2000). 'Degree operators and scope.' In Jackson B & Matthews T (eds.) *Proceedings of semantics and linguistic theory*, 10. Ithaca, NY: CLC Publications. 40–64.
- Hoeksema J (1984). 'Negative polarity and the comparative.' *Natural Language & Linguistic Theory* 1, 403–434.
- Kennedy C (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. New York: Garland Press.
- Kennedy C (2001). 'Polar opposition and the ontology of "degrees."' *Linguistics and Philosophy* 24, 33–70.
- Kennedy C (2002). 'Comparative deletion and optimality in syntax.' *Natural Language & Linguistic Theory* 20.3, 553–621.
- Klein E (1980). 'A semantics for positive and comparative adjectives.' *Linguistics and Philosophy* 4, 1–45.
- Klein E (1991). 'Comparatives.' In von Stechow A & Wunderlich D (eds.) *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin: Walter de Gruyter. 673–691.
- Larson R K (1988). 'Scope and comparatives.' *Linguistics and Philosophy* 11, 1–26.
- Rullmann H (1995). Maximality in the semantics of wh-constructions. Ph.D. diss., University of Massachusetts: Amherst.
- Sapir E (1944). 'Grading: A study in semantics.' *Philosophy of Science* 11, 93–116.
- Schwarzschild R & Wilkinson K (2002). 'Quantifiers in comparatives: A semantics of degree based on intervals.' *Natural Language Semantics* 10, 1–41.
- Seuren P A (1973). 'The comparative.' In Kiefer F & Ruwet N (eds.) *Generative grammar in Europe*. Dordrecht: Riedel. 528–564.
- Seuren P A (1978). 'The structure and selection of positive and negative gradable adjectives.' In Farkas D, Jacobsen W J & Todrys K (eds.) *Papers from the parasession on the lexicon*. Chicago: Chicago Linguistic Society. 336–346.
- Stassen L (1985). *Comparison and universal grammar*. Oxford: Basil Blackwell.
- von Stechow A (1984a). 'Comparing semantic theories of comparison.' *Journal of Semantics* 3, 1–77.
- von Stechow A (1984b). 'My reply to Cresswell's, Hellan's, Hoeksema's and Seuren's comments.' *Journal of Semantics* 3, 183–199.

Compositionality: Philosophical Aspects

F J Pelletier, Simon Fraser University,
Burnaby, BC, Canada

© 2006 Elsevier Ltd. All rights reserved.

There are three different but loosely related conceptions that are associated with the term 'compositionality' in the literature of philosophical and linguistic semantics.

One conception, taking its lead from the more literal sense of this technical term, concerns the manner of composition of objects in the world. In this sense, an object or type of object is compositional if it is identical with its parts when they are assembled in some specified way. A slogan for this notion of compositionality is: "An object is the sum of its parts." However, this is a slightly misleading slogan, because

it does not distinguish between two different types of objects made of the same parts but put together differently. This notion of compositionality is meta-physical in nature: it provides a characterization of the ontology of objects in the world, saying that they can all be described in terms of some basic atomic elements and their combinations. Along with this ontological feature often goes an epistemological feature: that one can know objects in the world by understanding what the atomic items are and the ways they can be assembled. Both the ontological and the epistemological aspects here are further associated with reductionism: the view that objects are “nothing more than” their parts. In this meaning of compositionality, the compositionists are often called ‘atomists,’ and anti-compositionists are called ‘holists’ or sometimes ‘wholists.’ These latter theorists deny that all objects can be described and known in terms of their parts and the arrangement of the parts – for instance, they might deny that a corporation, a nation, or a group is “nothing more than” the class of individuals making them up together with their relationships – and hence they are antireductionistic. They might also hold that there are emergent properties and gestalt properties that cannot be described and known in the way required by atomism. A slogan for these theories is: “The whole is more than the sum of its parts.”

In the field of semantics, whether semantics of natural language or of mental items, there is a somewhat different conception of compositionality in play. In this realm, it is meaning that is claimed to be compositional; but since meaning is always meaning of something, it is this other something that defines the parts and the whole, unlike the case of the first sort of compositionality. The slogan for this second conception of compositionality is: “The meaning of a whole is determined by the meaning of its parts and the way these parts are combined.” What we see here is that a **feature** of a whole (its meaning) is claimed to be **determined** by the similar feature in the parts of the whole, plus the mode of combination of these parts – unlike the case of the first type of compositionality, in which it was the whole itself that was alleged to be “nothing more than” its parts. In the second type of compositionality, the notions of ‘part’ and ‘whole’, as well as their mode of combination, are presupposed to be already defined in terms of an independent syntax (in the case of language) or an independent mental economy (in the case of concepts). So the realm of syntax or mental economy is presupposed to be compositional in the first sense, and the issue is whether the property of meaning that is associated with the parts and wholes will likewise compose. Since the second conception assumes that

the first conception applies to the background syntax, this second conception presupposes basic or primitive meanings for the atomic (syntactic or mental) parts out of which all other (syntactic or mental) items are composed. (Once this second notion of compositionality is acknowledged, where there is a presupposed part-whole structure and it is then asked whether a feature of the whole is somehow determined by the similar features in the parts, one can see questions of compositionality arising in many fields, not just in semantics. For example, one might wonder whether the intrinsic value of an action is determined by the values of the parts of the action and the way the parts are ordered. One might wonder whether the beauty of a whole is determined by the beauty of its parts and the way the parts are combined. One might wonder whether the duties and obligations of a corporation or a society are determined by those of its members and the way these members fit together to form the corporation or society.)

Obviously, whether semantic compositionality is true or false depends upon the presupposed syntax or mental economy, the conception of meaning under consideration, and what is meant by the phrase “is determined by.” Indeed, many theorists have thought that this indeterminism inherent in semantic compositionality shows that its truth or falsity is merely “a methodological matter.” For a small alteration in the underlying syntax or mental economy might make a given semantics become non-compositional; a slight change in the assumed notion of ‘determination’ might make it become compositional again; an inclusion or exclusion of some property as “being semantic meaning” (as opposed, say, to “being pragmatics”) makes it become non-compositional again; and there might be no reason to make these changes other than to keep or deny compositionality.

The most popular explanation of “is determined by” in the semantic compositionist’s slogan is that it means ‘is a (mathematical) function of’; so the slogan becomes: “The meaning of a complex syntactic unit is a (mathematical) function of the meanings of its syntactic parts and the way in which they are syntactically combined.” But according to some, this notion allows too much: it is claimed that if no constraints are put upon the function, nearly any meanings of parts and syntactic combination can be compositionally related to the meaning of a whole. Some theorists would want to make the function be natural or systematic (and so on), without saying much about what, exactly and in the abstract, would make a function be natural or systematic. More usual is to be given examples of what sort of mathematical function should be ruled out. Consider the idea that an adjective like *red* means something

different depending on what noun it modifies. For example, according to this view, *red wine* vs. *red rose* vs. *red hair* vs. *red skin* vs. *red grapefruit* all employ a different meaning of *red*. And then compositionality is false, because these phrases are all constructed by the same syntactic rule and yet the meaning of *red* changes as a result of some syntactic item (viz., the noun being modified) that is not a part of the lexical item *red*. But a defender of compositionality could respond that the meaning of *red* is constant throughout, by being disjunctive (“when modifying *wine* it means r_1 ; when modifying *hair* it means r_2 ; etc.”). This is a perfectly good mathematical function and would obviously yield the right meanings of wholes if there were enough disjuncts. Those who object to the mathematical notion of function in the definition of compositionality might claim here that disjunctive functions are “not natural.”

The notion opposed to semantic compositionality is ‘semantic holism’. However, this notion means different things to different theorists, and it is not always just taken to mean merely that there is no mathematical function that will generate the required meanings. For example, some people call semantic holism the view that “words have meaning only in the context of a sentence” or that no word or other syntactic unit (including sentences, paragraphs, and discourses) has meaning in itself, but only in the setting of an entire theory or worldview or form of life. Others take semantic holism to be that the meaning of a syntactically defined item is determined not only by the meanings of its syntactic parts and their syntactic combination but also by the nonlinguistic context in which the utterance is made. (For example, it might be thought that the meaning of *There is no money* depends on who is speaking, whether the audience knows which business deal is being discussed, and so forth.) And still other holists, not necessarily wanting to bring these nonlinguistic items into meaning, nonetheless might hold that there are cases where the meaning of a syntactically complex item depends on meanings of linguistic items that are not syntactic parts of the complex. (For example, in *The first man landed on the moon in 1969*, we cannot take the meaning of *the first man* and combine it with *landed on the moon in 1969* to get the right meaning, for there is no sense in which the sentence really is talking about *the first man*. Rather, the relevant meaning of the subject term is that of *the first man who landed on the moon*. But to obtain that meaning, we need to get information from the verb phrase. Hence, to get the meaning of the subject term we need information of items that are not syntactic parts of the subject term.)

A third conception for (semantic) compositionality is less definite than the preceding, and comes through

considerations that might be called ‘the magic of language’. A set of closely related considerations have been pointed at in various times in the history of philosophy, both Western and Indian:

- We can **understand** an infinite number of novel sentences, so long as they employ words we already understand. We understand sentences and combinations that we have never encountered.
- We can **create** new sentences that we have never heard or used before, and we know that they are appropriate to the situation in which we use them.
- We are finite creatures who are exposed to a finite amount of information concerning our language. Nonetheless, we **learn** a system that is capable of infinite expression.

These considerations all point to the same features: (1) that language is something special (infinite, novel, creative, or whatever) and (2) that people manage to use/learn/understand language despite their finite nature. It is natural to see compositionality as an explanation of this ability – people have a finite stock of atomic items whose meanings are learned primitively, and there is a finite number of rules of combination whose effect on meaning are learned. But given that the rules are recursive in nature, this allows for an infinite number of sentences whose meanings are finitely knowable. (The opening paragraph of Frege [1923] is often taken to be an endorsement of this argument for compositionality, but it is a matter of scholarly dispute as to whether or not Frege actually believed in semantic compositionality. See Pelletier, 2001 and Janssen, 2001 for discussion and further references.)

This third conception of (semantic) compositionality is a ‘functional’ one and thus less definite than the preceding two. It amounts to saying that compositionality is *whatever* accounts for the magic of language. It might be the second conception of compositionality, with its mathematical functions, that will do the trick, or it might be some other, more exotic type of function. Or it may be some function that operates on items that are not necessarily syntactic subparts of the expression to be evaluated, and somehow thereby brings in information from context (of both linguistic and nonlinguistic varieties).

The magic of language considerations are the only arguments in favor of compositionality that do not seem merely to turn on such methodological considerations as the aesthetics of the syntax-semantics interface. However, it should be noted that they are not conclusive in relation to compositionality-as-mathematical-function. The second notion of compositionality does not guarantee the magic, nor does

non-compositionality in this second notion necessarily deny the magic. For it might be that the meaning of every syntactic whole is a function of the meanings of its parts and its syntactic mode of combination, but if these functions are not computable functions, then the language cannot be learned/used/understood in the way required by the magic. On the other hand, even if there is no function defined solely by the meanings of the parts and their modes of combination that will yield the meanings of the wholes, it could nonetheless be true that these meanings are computable in some other way ... and then the magic would still be there. (An example of this possibility is Pelletier's 1994/2004 'semantic groundedness'.)

Considerations Against Semantic Compositionality

The linguistic semantics literature is rife with demonstrations of how some linguistic phenomenon can or cannot be given a compositional description. It often seems that these works would more accurately be described as demonstrating how a phenomenon can or cannot be given a compositional description **employing some particular syntactic-semantic device or within some specific syntactic-semantic theory**. There are, however, three more general arguments that have been presented against semantic compositionality. The first is an argument from (nonlinguistic) context, of the sort mentioned above, where it is claimed that the meaning of a sentence in a context just cannot be derived from the meanings of the words and their combinations. In evaluating this argument, scholars need to distinguish between (what might be called) 'literal meaning' and 'occasion meaning'. The former is thought of as the meaning-in-language, while the latter is thought of as the meaning-in-a-context. If there is such a distinction, then there will be two principles of semantic compositionality – one for each type of meaning. And it is not so clear that either of them is overturned by considerations of context. The only casualty would be a mixed principle that no one believes, i.e., that the occasion meaning of a complex expression is a mathematical function of the literal meanings of its parts and their manner of combination.

The second general argument against compositionality comes from the existence of synonymy and Mates-like (Mates, 1950) considerations. Given that there is synonymy, so that x_1 and x_2 mean the same, then there are two sentences, S_1 and S_2 , that differ only in that one contains x_1 while the other contains x_2 . Given compositionality, it follows that S_1 and S_2 are synonymous too; and by compositionality again, it follows that *Mary believes S_1* and *Mary believes S_2*

are synonymous. But for any such S_1 and S_2 , it can be the case that the former is true, while the latter is false. However, it cannot be the case that, of two synonymous sentences, one is true and the other false. Hence, either there is no synonymy or else compositionality is wrong. And the existence of synonymy is more secure than that of compositionality.

The third general argument comes from the existence of ambiguity. If compositionality implies that the meaning of a whole is a mathematical function of the meanings of its parts (and combination), then there cannot be any ambiguity of the sort where **one and the same item** has two or more meanings, for that would deny that it was a function that computed meaning. As with synonymy, one could of course deny the existence of ambiguity; but most theorists find that this is too lavish a position to take. So it is usually admitted by compositionlists that individual words can be ambiguous; therefore, sentences using these ambiguous words may also be ambiguous (but the ambiguities are always traceable to the ambiguity of the words). Also, it is pointed out that strings of words such as *Visiting professors can be fun* are ambiguous (is it the professors or the activity of visiting the professors that can be fun?), but this ambiguity is traceable to the fact that the words are put together in different ways – that is, there are different structural descriptions that can be associated with this string of words. Hence, this ambiguity is not a challenge to compositionality. However, Pelletier (1994/2004) points to a number of examples that seem neither to have ambiguous words nor to have different structural descriptions but which are nonetheless ambiguous. For example: *When Alice rode a bicycle, she went to school*. This seems to have but one syntactic analysis within any particular theory, but its meaning is ambiguous: *On those occasions where Alice rode a bicycle, she took it to school* vs. *Back in the days when Alice was a bicyclist, she was a student*.

Formal Considerations

There have been a number of works concerned with the question of whether compositionality is a nonempirical issue on the grounds of certain formal features that are required by compositionality. A review article that surveys this work is Westerståhl (1998). More recent work on formal features of compositional semantics is in the important work of Hodges (2001) and material based on this.

History

Although the general principle of compositionality seems to have been around for some time, as mentioned earlier, it is not clear when the term

'compositionality' came into the linguistic semantics literature (unlike 'holism,' which was introduced by Smuts, 1926). 'Compositionality' is used by Katz (1973) and Thomason (1974).

See also: Context Principle; Holism, Semantic and Epistemic; Representation in Language and Mind; Systematicity.

Bibliography

- Davidson D (1965). 'Theories of meaning and learnable languages.' In Bar-Hillel Y (ed.) *Logic, methodology and philosophy of science*. Amsterdam: North Holland. 383–394.
- Dever J (in press). 'Compositionality.' In Lepore E & Smith B (eds.) *Oxford handbook of the philosophy of language*. Oxford: Oxford University Press.
- Fodor J & Lepore E (1992). *Holism: a shopper's guide*. Oxford: Blackwell.
- Fodor J & Lepore E (2002). *The compositionality papers*. New York: Oxford University Press.
- Frege G (1923/1963). 'Compound thoughts.' Stoothoff R (trans.). *Mind* 72, 1–17.
- Hodges W (2001). 'Formal features of compositionality.' *Journal of Logic, Language and Information* 10, 7–28.
- Janssen T (1997). 'Compositionality.' In van Benthem J & ter Meulen A (eds.) *Handbook of logic and language*. Amsterdam: Elsevier. 417–473.
- Janssen T (2001). 'Frege, contextuality and compositionality.' *Journal of Logic, Language and Information* 10, 115–136.
- Kamp H & Partee B (1995). 'Prototype theory and compositionality.' *Cognition* 57, 129–191.
- Katz J (1973). 'Compositionality, idiomaticity, and lexical substitution.' In Anderson S & Kiparsky P (eds.) *A festschrift for Morris Halle*. New York: Holt, Rinehart, and Winston. 357–376.
- Mates B (1950). 'Synonymity.' *California University Publications in Philosophy* 25. Reprinted in Linsky L (1952). *Semantics and the philosophy of language*. Urbana: University of Illinois Press. 111–136.
- Pagin P (1997). 'Is compositionality compatible with holism?' *Mind and Language* 12, 11–23.
- Partee B (1984). 'Compositionality.' In Landman F & Veltman F (eds.) *Varieties of formal semantics*. Dordrecht: Foris. 281–311.
- Partee B (2003). *Compositionality in formal semantics: selected papers by Barbara H. Partee*. Oxford: Blackwell.
- Pelletier F J (1994/2004). 'The principle of semantic compositionality.' *Topoi* 13, 11–24. [Reprinted with new appendices in Davis S & Gillon B (eds.) *Semantics: a reader*. New York: Oxford University Press. 133–158.]
- Pelletier F J (2001). 'Did Frege believe Frege's principle?' *Journal of Logic, Language and Information* 10, 87–114.
- Smuts J (1926). *Holism and evolution*. London: Macmillan.
- Szabo Z (2000). *Problems of compositionality*. New York: Garland.
- Thomason R (1974). 'Introduction.' In Thomason R (ed.) *Formal philosophy: selected papers of Richard Montague*. New Haven: Yale University Press. 1–69.
- Westerståhl D (1998). 'On mathematical proofs of the vacuity of compositionality.' *Linguistics and Philosophy* 21, 635–643.
- Westerståhl D (2002). 'Idioms and compositionality.' In Barker-Plummer D, Beaver D, van Benthem J & Scottodi Luzio P (eds.) *Words, proofs, and diagrams*. Stanford: CSLI Publications. 241–271.

Compositionality: Semantic Aspects

G Sandu and P Salo, University of Helsinki, Helsinki, Finland

© 2006 Elsevier Ltd. All rights reserved.

According to the principle of compositionality, the meaning of a complex expression depends only on the meanings of its constituents and on the way these constituents have been put together. The kind of dependence involved here is usually a functional one.

Principle of Compositionality (PC): The meaning of complex expression is a function of the meanings of its constituents and of the rule by which they were combined.

PC is rather vague unless one specifies the meanings of 'is a function of' and 'meaning(s)', something that is easier said than done. A more rigorous formulation

of these notions is possible for formal languages and is due to Richard Montague.

Montague (1974) defined compositionality as the requirement of the existence of a homomorphism between syntax and semantics, both to be understood as 'structures' in the mathematical sense. To keep technicalities down to a minimum, Montague's requirement of a compositional interpretation was that for each syntactic operation 'O' that applies to n expressions e_1, \dots, e_n in order to form the complex expression ' $O(e_1, \dots, e_n)$ ', the interpretation of the complex expression ' $O_i(e_1, \dots, e_n)$ ' is the result of the application of the semantic operation ' C_i ', which is the interpretation of ' O_i ' to the interpretations m_1, \dots, m_n of ' e_1, \dots, e_n ', respectively. In other words, the interpretation of ' $O_i(e_1, \dots, e_n)$ ' is $C_i(m_1, \dots, m_n)$.

An immediate consequence of PC is the ‘Substitutivity Condition’: Substituting a constituent with its synonym in a given expression does not change the meaning of the resulting expression. Thus, PC is violated if a complex expression has meaning but some of its component expressions do not (the Domain Condition) or if the Substitutivity Condition fails.

As one can see, PC is by itself rather weak, and so it comes as no surprise that in the case of formal languages, one can always devise a trivial compositional interpretation by assigning some arbitrary entities to the primitive expressions of the language and then associating arbitrarily the syntactic operations of the language with corresponding operations on the domain of those entities. This way of implementing the principle can hardly be of any interest, although it has led some philosophers and logicians to claim that PC is methodologically empty.

A slightly more interesting case is the one in which one has an intended semantic interpretation in mind, that is, an interpretation with an intended domain of entities for the primitive expressions of the language to be mapped into, and a class of intended operations to serve as the appropriate interpretations of the syntactic operations of the language. A case in point is Horwich’s (1998) interpretation. His formal language was intended to serve as a regimentation for a fragment of English that contains proper names (‘John,’ ‘Peter,’ etc.), common nouns (‘dogs,’ ‘cows,’ etc.), and verb phrases (‘talks,’ ‘walks,’ ‘bark,’ etc.) as primitive expressions together with grammatical operations on them. For simplicity, let us assume predication is such a grammatical operation marked in this case by an empty space. Thus the syntax contains clauses of the form:

If ‘*n*’ is a proper name and ‘*v*’ is a verb phrase, then ‘*n v*’ is a complex expression.

The intended semantic interpretation consists of a domain of entities that serve as the intended meanings of the proper names and verbs phrases (whatever they are; they are marked by capitals), together with an operation – say, *P* – that interprets the grammatical operation of predication (whatever that is). The only thing one needs to worry about in this case is to see to it that the operation of predication is defined for the entities mentioned above. The relevant semantic clauses now have this form:

The interpretation of ‘*n v*’ is the result of the application of *P* to the entities assigned to ‘*n*’ and ‘*v*’, respectively.

Thus, the interpretation of the sentence ‘John talks’ is the result of the application of *P* to TALKS to JOHN. This interpretation is trivially compositional in that the interpretation of every compound ‘*n v*’ has been defined as the result of the application of the

operation assigned to the syntactic operation of concatenation to the interpretations of ‘*n*’ and ‘*v*’, respectively. The more challenging cases for PC are those in which one has an intended interpretation for the complex expressions and would like to find a compositional interpretation that agrees with it. In contrast to the previous case, the meanings of the complex entities are not any longer defined but are given at the outset. We have here a typical combination of PC with the *Context Principle* (CP): An expression has a meaning only in the context in which it occurs. The combination was largely explored in the work of Gottlob Frege and in Donald Davidson’s theory of meaning, which assumed the form of a theory of truth. Davidson took whole sentences to be the meaning-carrying units in language, and truth to be a primitive, undefinable semantic property that is best understood. Truth being undefinable, the strategy applied above, which ensured a trivial implementation of PC, is no longer available. Instead, PC acquires the status of a methodological constraint on an empirical theory of truth for the target language: the division of a sentence into parts and their association with appropriate semantic entities in a compositional theory becomes a theoretical business that has no other role except to show how they contribute to the computation of the truth of the sentences of the target language in which they occur.

The literature on formal semantics for natural language has plenty of cases of the application of the Context Principle. We consider just two examples. In *game-theoretical semantics* (GTS), one starts with a standard first-order language and defines truth only for the sentences of that language. The truth of every such sentence (in a prenex normal form) is defined via a second-order sentence, known as its Skolem form. This interpretation is clearly not compositional, since it violates the Domain Condition. One can now ask whether there is a compositional interpretation that agrees with the given game-theoretical interpretation of sentences. It is known that the answer is positive, but only assuming certain nontrivial mathematical principles (the Axiom of Choice).

The second example concerns Dynamic Predicate Logic. The starting point is the same language as in GTS – that is, a standard first-order language – but we now want a compositional interpretation in which, e.g., an existential quantifier occurring in the antecedent of a conditional binds a free variable occurring in the consequence of the conditional and in addition has the force of an universal quantifier. There is a compositional interpretation that has the required property, that of Dynamic Predicate Logic (Groenendijk and Stokhoff, 1991).

From a technical point of view, the situation described in the two examples may be depicted as an extension problem (Hodges, 1998). One starts with an intended interpretation *I*, which either (a) fixes only the interpretation of certain complex expressions (e.g., sentences) or (b) puts some relational constraints on the interpretation of complex expressions. One then wants to find a compositional interpretation *I'* that agrees with the independently understood interpretation *I*. Hodges's Extension Theorem solves case (a). It shows that any partial interpretation for a grammar can be extended to a total compositional interpretation. This shows that the combination of PC with CP (in its form [a]) is trivially satisfiable. The more interesting cases are those falling under (b). This is the situation that typically arises in the case of empirical linguistics where the intended interpretation is supposed to be motivated by empirical argument.

As an illustration, consider the much-discussed 'pet fish' problem. There is some empirical evidence to the effect that the meanings of concept words are prototypes. A prototype is either a good exemplar of the category or a statistical average of all or some instances of the category (Smith and Medin, 1981). A given instance *x* is then categorized as *X* if *x* resembles the prototype of *X* more than any other prototype. Given two expressions *X* (e.g., 'pet') and *Y* ('fish'), one asks whether there is an interpretation that assigns to the complex concept word *XY* ('pet fish') a prototype that is the composition of the prototype assigned to *X* and the prototype assigned to *Y*. One also wants the meaning function to satisfy certain basic properties that are required for explanatory purposes; e.g., it should be the case that if *x* is *XY*, it must also be *X* and *Y*. We thus want every *x* to resemble the prototype of *XY* no less than it resembles the prototypes of *X* and *Y*. It has been argued that there is no such interpretation, that is, there is no operation of composition that yields a prototype as the interpretation of *XY* with the desired properties when applied to the two prototypes that are the interpretations of *X* and *Y* respectively (Fodor, 1998; Osherson and Smith, 1981).

The moral to be drawn from all this should have been somehow anticipated from our discussion of formal languages. When the intended interpretation puts constraints only on the meanings of primitive expressions and on the operations governing them, PC follows rather trivially, provided the semantic entities of complex expressions are not constrained in any way.

When the intended interpretation concerns only the meanings of complex expressions, Hodges's extension theorem shows that a compositional semantics can still be found, at least in some cases, provided that one does not constrain the meanings of the primitive

expressions or syntactical operations on them. In natural language, however, the situation is hardly so simple, as one meets constraints at every level. It is no wonder, then, that Fodor and Lepore (2002) argued that most theories of concepts or mental architecture in cognitive science are in contradiction with PC. The case of prototype semantics was only one example, but the same considerations apply to the theory that the meaning of a word is its use or the criteria for its application, etc.

PC is often defended as the best explanation of the empirical phenomenon of systematicity: Any competent speaker of a given language who has in his repertoire the complex expressions *P*, *R*, and *Q* has also in his repertoire the complex expressions in which *P*, *R*, and *Q* are permuted (provided they are grammatical). For instance, anybody who understands the sentence 'Mary loves John' also understands the sentence 'John loves Mary'. Fodor and his collaborators argued extensively that PC is the best explanation of the systematicity of language, but this is an issue that will not be tackled here (cf. Fodor and Pylyshyn, 1988; Fodor, 2001; Fodor and Lepore, 2002; Fodor, 2003; Aizawa, 2002).

PC should not be confused with the principles of productivity or generativity of language, which require that the expressions of a language be generated from a finite set of basic expressions and syntactical rules. Although it presupposes that the language under interpretation has a certain syntactic structure, PC does not take a stand on how that structure should be specified (phrase structure rules, derivational histories, etc.), as long as it is given a compositional interpretation.

See also: Compositionality: Philosophical Aspects; Context Principle; Game-Theoretical Semantics; Systematicity.

Bibliography

- Aizawa K (2002). *The systematicity argument*. Amsterdam: Kluwer.
- Bloom P (1994). 'Generativity within language and other domains.' *Cognition* 51(2), 177–189.
- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton.
- Fodor J A (1998). *Concepts: where cognitive science went wrong*. Oxford: Clarendon Press.
- Fodor J A (2001). 'Language, thought and compositionality.' *Mind and Language* 16(1), 1–15.
- Fodor J A (2003). *Hume variations*. Oxford: Oxford University Press.
- Fodor J A & Lepore E (2002). *The compositionality papers*. Oxford: Clarendon Press.
- Fodor J A & Pylyshyn Z (1988). 'Connectionism and cognitive architecture: a critical analysis.' *Cognition* 28, 3–71.

- Groenendijk J & Stokhoff M (1991). 'Dynamic predicate logic.' *Linguistics and Philosophy* 14, 39–100.
- Hintikka J & Kulas J (1983). *The game of language*. Dordrecht: Reidel.
- Hodges W (1998). 'Compositionality is not the problem.' *Logic and Philosophy* 6, 7–33.
- Horwich P (1998). *Meaning*. Oxford: Clarendon Press.
- Janssen T M V (1997). 'Compositionality.' In van Benthem J & Meulen A T (eds.) *Handbook of logic and language*. Amsterdam: Elsevier. 417–473.
- McLaughlin B (1993). 'The classicism/connectionism battle to win souls.' *Philosophical Studies* 70, 45–72.
- Montague R (1974). *Formal philosophy: selected papers of Richard Montague*. New Haven: Yale University Press.
- Osherson D N & Smith E E (1981). 'On the adequacy of prototype theory as a theory of concepts.' *Cognition* 9, 35–58.
- Pelletier F J (1994). 'The principle of semantic compositionality.' *Topoi* 13, 11–24.
- Rips L J (1995). 'The current status of research on concept combination.' *Mind and Language* 10(1/2), 72–104.
- Smith E E & Medin D L (1981). *Categories and concepts*. Cambridge: Harvard University Press.
- Smolensky P (1987). 'The constituent structure of mental states: a reply to Fodor and Pylyshyn.' *Southern Journal of Philosophy* 26, 137–160.
- Zadrozny W (1994). 'From compositional to systematic semantics.' *Linguistics and Philosophy* 17, 329–342.

Concepts

E Margolis, Rice University, Houston, TX, USA

S Laurence, University of Sheffield, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

In cognitive science, concepts are generally understood to be structured mental representations with subpropositional content. The concept CHAIR, for example, is a mental representation with the content *chair*. It is implicated in thoughts about chairs and is accessed in categorization processes that function to determine whether something is a chair. Theories of concepts are directed to explaining, among other things, the character of these processes and the structure of the representations involved. Related to this is the project of explaining what conceptual content is and how concepts come to have their content. In the study of conceptual structure, four broad approaches should be distinguished: (1) the classical theory, (2) probabilistic theories, (3) the theory-theory, and (4) conceptual atomism. For recent overviews of theories of concepts, see Margolis and Laurence (1999) and Murphy (2002).

The Classical Theory

According to the classical theory, concepts have definitional structure. A concept's constituents encode conditions that are individually necessary and jointly sufficient for its application. A standard illustration of the theory is the concept BACHELOR, which is claimed to be composed of the representations UNMARRIED, ADULT, and MALE. Each of these is supposed to specify a condition that something must meet in order to be a bachelor and, if anything meets them all, it is a bachelor.

The classical theory has always been an enormously attractive theory. Many theorists find it to be intuitively plausible that our concepts are definable. In addition, the theory brings with it a natural and compelling model of how concepts are learned. They are learned by assembling them from their constituents. The classical theory also offers a straightforward account of categorization. Something is deemed to fall under a concept just in case it satisfies each and every condition that the concept's constituents encode. Finally, the theory appeals to the very same resources to explain the referential properties of a concept. A concept refers to those things that have each and every feature specified by its constituents. Of course, all of these explanations depend upon there being a separate treatment of the primitive (i.e., unstructured) representations that ultimately make up the concepts we possess. But the classical theory supposes that a separate treatment can be given, perhaps one that grounds all of our concepts in perceptual primitives in accordance with traditional empiricist models of the mind.

The classical theory has come under considerable pressure in the last thirty years or so. In philosophy, the classical theory has been subjected to a number of criticisms but perhaps the most fundamental is that attempts to provide definitions for concepts have had a poor track record. There are few – if any – examples of uncontroversial definitional analyses. The problem isn't just confined to philosophically interesting concepts (e.g., JUSTICE) but extends to concepts of the most ordinary kind, such as GAME, PAINT, and even BACHELOR (Wittgenstein, 1953; Fodor *et al.*, 1980). What's more, Quine's (1951) influential critique of the analytic-synthetic distinction has led

many philosophers to suppose that the problem with giving definitions is insurmountable.

For psychologists, the main objection to the classical theory has been that it appears to be at odds with what are known as ‘typicality effects.’ Typicality effects include a broad range of phenomena centered around the fact that certain exemplars are taken to be more representative or typical (Rosch and Mervis, 1975; Rosch, 1978). For instance, apples are judged to be more typical than plums with respect to the category of fruit, and subjects are quicker to judge that apples are a kind of fruit than to judge that plums are and make fewer errors in forming such judgments. Though not strictly inconsistent with these findings, the classical theory does nothing to explain them.

Probabilistic Theories

In response to the failings of the classical theory, Eleanor Rosch and others began exploring the possibility that concepts have a structure that is described as graded, probabilistic, or similarity-based (Smith and Medin, 1981). The difference between these approaches and the classical theory is that the constituents of a concept are no longer assumed to express features that its members have by definition. Instead, they are supposed to express features that its members tend to have. For example, a standard treatment for the concept BIRD incorporates constituents picking out the features *has wings*, *flies*, *eats worms*, etc., but probabilistic theories don’t require all of these features to be possessed by something to count as a bird. Instead, something falls under the concept when it satisfies a sufficient (weighted) number of them (or on some accounts, something falls under the concept to a degree corresponding to how many are satisfied; then nothing is a bird absolutely but only a bird to degree n).

Like the classical theory, probabilistic theories explain concept learning as a process where a concept is assembled from its constituents. And like the classical theory, probabilistic theories offer a unified treatment of reference and categorization. A concept refers to those things that satisfy enough of the features it encodes, and something is judged to fall under a concept when it satisfies enough of them as well. Categorization, on this account, is often described as a similarity comparison process. An item is categorized as belonging to a given category when the representations for each are deemed sufficiently similar, where this may be measured in terms of the number of constituents that they share.

One advantage of probabilistic theories is that a commitment to probabilistic structure may explain why definitions are so hard to come by. More important, however, is the way that probabilistic structure

readily accommodates and explains typicality effects. This is achieved by maintaining that typicality, like categorization, is a similarity comparison process. On this model, the reason apples are judged to be more typical than plums is that the concept APPLE shares more of its constituents with FRUIT. Likewise, this is why apples are judged to be a kind of fruit faster than plums are.

Probabilistic theories continue to enjoy widespread support in cognitive science, but they aren’t without their own problems. One concern is that many concepts appear to lack probabilistic structure, especially concepts that correspond to phrases as opposed to words. For example, Fodor (1981), (1998) notes that while GRANDMOTHER may have probabilistic structure (encoding the features *gray-haired*, *old*, *kind*, etc.), there is no such structure for GRANDMOTHERS MOST OF WHOSE GRANDCHILDREN ARE MARRIED TO DENTISTS. Fodor also challenges probabilistic theories on the grounds that even when phrasal concepts do have probabilistic structure, their structure doesn’t appear to be compositionally determined. This is a problem, since it’s the compositionality of the conceptual system that explains the productivity of thought, viz., the fact that there is no upper bound on the number of distinct thoughts that humans can entertain. Fodor points out that the probabilistic structure associated with PET FISH encodes features (*colorful*, *tiny*, *lives in a bowl*, etc.) that aren’t drawn from the probabilistic structures associated with PET (*furry*, *cuddly*, etc.) and FISH (*gray*, *lives in the ocean*, etc.).

Another common criticism of probabilistic theories is that they leave out too much. They don’t sufficiently incorporate the causal information that people appeal to in categorization and don’t do justice to the fact that reflective categorization isn’t always based on similarity (Murphy and Medin, 1985; Keil, 1989; Rips, 1989). For example, when time is short and when given little information about two animals apart from the fact that they look alike, people may judge that they are both members of the same category. But when asked for a more thoughtful answer about whether, for example, a dog that is surgically altered to look like a raccoon is a dog or a raccoon, the answer for most of us – and even for children – is that it remains a dog (see Gelman, 2003, for an overview of related literature).

The Theory-Theory

The theory-theory is largely a reaction to the last problem associated with probabilistic theories. It explains categorization, particularly reflective categorization, as a process of causal-explanatory reasoning. On this approach, conceptual structure is

a matter of how a concept is related to other concepts in relatively stable causal-explanatory frameworks. The designation ‘theory-theory’ sometimes implies little more than this. For some psychologists, it is meant to indicate that the explanatory frameworks are comparable to explicit scientific theories and that the mechanisms for acquiring them are identical with the cognitive mechanisms that underlie scientific reasoning. On this more extreme version of the theory-theory, conceptual development is likened to radical theory change in science (Carey, 1985; Gopnik and Meltzoff, 1997).

Many objections to the theory-theory are directed to its more extreme forms, particularly the commitment about conceptual development. The claim that infants are like little scientists has generated a great deal of criticism (e.g., Segal, 1996; Stich and Nichols, 1998). One objection focuses on particular examples, especially of concepts that are fundamental to human cognition (e.g., OBJECT, AGENT, and BELIEF). Although theory-theorists often cite these as examples where substantial conceptual change occurs – change that is supposed to illustrate the theory-theory’s model of cognitive development – others would argue that these are innate concepts that remain invariant in important respects throughout development (e.g., Leslie, 1994). A more basic objection to the theory-theory is that the appeal to causal-explanatory reasoning is minimally informative. It may be true that categorization is somewhat like scientific reasoning, but scientific reasoning is itself in need of a great deal of clarification. The result is that the model of categorization is extremely sketchy and somewhat mysterious.

A third objection to the theory-theory, one that has been especially influential in philosophy, is that it makes it difficult to maintain that different people have the same concepts. This objection is directed to versions of the theory-theory that are especially lenient in what counts as a theory. On these versions, just about any belief or inferential disposition associated with a concept is part of a ‘theory.’ The problem with this approach, however, is that people are bound to have different beliefs than one another and hence different theories. But since a concept’s identity and content are supposed to be a matter of its role in one’s mental theories, people will be unable to share concepts (Fodor and Lepore, 1992).

Conceptual Atomism

The last of the four theories of conceptual structure is that lexical concepts – word-sized concepts – have no structure at all (Fodor, 1998; Millikan, 2000).

Concepts such as BIRD, CHAIR, NUMBER, and RUN are all primitives. Of course, conceptual atomism needs an account of how these primitive concepts are to be distinguished from one another and how their contents are fixed. A standard approach is to appeal to the mind-world causal relations between a concept and the object or property it refers to.

Conceptual atomism is motivated in light of the problems with other theories, especially the problem of providing definitions (the classical theory), the problem of compositionality (probabilistic theories), and the problem of shared concepts (the theory-theory). If concepts lack structure, then it is no surprise that we have difficulty providing definitions for them. Also, it doesn’t matter that probabilistic structure doesn’t compose, since **complex** concepts can still be composed on the basis of atomic constituents. And sharing a concept is no longer a challenge. It isn’t a matter of having the same beliefs so much as having representations that stand in the same mind-world causal relations.

Conceptual atomism is sometimes rejected outright on the grounds that unstructured concepts can’t be learned and hence that atomism implies an untenably strong form of concept nativism (see **Innate Ideas**). The main concern with conceptual atomism, however, is that without structure, there is nothing to explain how concepts are implicated in categorization and other psychological processes. Nonetheless, atomists see this as an advantage rather than a problem, maintaining that people can have the same concept despite widely varying psychological dispositions. For this reason, the structures that are accessed in categorization and other psychological processes are said to be associated with a concept but not constitutive of it.

See also: Analytic/Synthetic, Necessary/Contingent, and *a Priori/a Posteriori*; Distinction; Causal Theories of Reference and Meaning; Compositionality: Philosophical Aspects; Definitions: Uses and Varieties of; Holism, Semantic and Epistemic; Innate Ideas; Mentalese; Representation in Language and Mind.

Bibliography

- Carey S (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Fodor J A (1981). ‘The present status of the innateness controversy.’ In *his Representations: philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press. 257–316.
- Fodor J A (1998). *Concepts: where cognitive science went wrong*. New York: Oxford University Press.

- Fodor J A, Garrett M, Walker E & Parkes C (1980). 'Against definitions.' *Cognition* 8, 263–367.
- Fodor J A & Lepore E (1992). *Holism: A shopper's guide*. Cambridge, MA: Basil Blackwell.
- Gelman S (2003). *The essential child*. New York: Oxford University Press.
- Gopnik A & Meltzoff A (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Keil F (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Leslie A (1994). 'ToMM, ToBy, and agency: core architecture and domain specificity.' In Hirshfeld L & Gelman S (eds.) *Mapping the mind: domain specificity in cognition and culture*. New York: Cambridge University Press. 119–148.
- Margolis E & Laurence S (1999). *Concepts: core readings*. Cambridge, MA: MIT Press.
- Millikan R (2000). *On clear and confused ideas*. New York: Cambridge University Press.
- Murphy G (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Murphy G & Medin D (1985). 'The role of theories in conceptual coherence.' *Psychological Review* 92(3), 289–316.
- Quine W (1951). 'Two dogmas of empiricism.' In *his From a logical point of view: nine logico-philosophical essays*. Cambridge, MA: Harvard University Press. 20–46.
- Rips L (1989). 'Similarity, typicality, and Categorization.' In Vosniadou S & Ortony A (eds.) *Similarity and analogical reasoning*. New York: Cambridge University Press. 21–59.
- Rosch E (1978). 'Principles of categorization.' In Rosch E & Lloyd B (eds.) *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates. 27–48.
- Rosch E & Mervis C (1975). 'Family resemblances: studies in the internal structure of categories.' *Cognitive Psychology* 7, 573–605.
- Segal G (1996). 'The modularity of theory of mind.' In Carruthers P & Smith P (eds.) *Theories of theories of mind*. Cambridge: Cambridge University Press. 141–158.
- Smith E & Medin D (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Stich S & Nichols S (1998). 'Theory-theory to the max.' *Mind and Language* 13(3), 421–449.
- Wittgenstein L (1953). *Philosophical investigations*. Anscombe (trans.). Oxford: Blackwell.

Conditionals

S Kaufmann, Northwestern University, Evanston, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Form and Meaning

Conditionals are complex sentences built up from two constituent clauses, called the *antecedent* and the *consequent*; alternatively, the terms *protasis* and *apodosis* are found in the linguistic literature. English conditionals are typically of the form *if A, (then) B*, where *A* and *B* are the antecedent and consequent, respectively. Some examples are given in (1).

- (1a) If the sun comes out, Sue will go on a hike.
- (1b) If the sun came out, Sue went on a hike.
- (1c) If the sun had come out, Sue would have gone on a hike.

In the linguistic and philosophical literature, a distinction is commonly drawn between *indicative* conditionals, such as (1a) and (1b), and *subjunctive* or *counterfactual* conditionals, like (1c). This classification is not uncontroversial: some authors would draw the major dividing line between (1a) and (1c) on the one hand and (1b) on the other. However, we

adopt the standard classification and focus on indicative conditionals (see also **Counterfactuals**). The class of indicatives may be further divided into *predictive* and *nonpredictive* conditionals, illustrated in (1a) and (1b), respectively. Despite subtle differences, these share a common semantic core and have similar logical properties. We do not distinguish between them in this discussion.

In general, *if A, B* asserts that *B* follows from, or is a consequence of *A*, without asserting either *A* or *B*. Often the relation in question is causal (*A* causes *B*) or inferential (*B* is inferable from *A*). Other uses include the statement that *B* is *relevant* if *A* is true (2a), conditional speech acts (2b), and metalinguistic comments on the consequent (2c).

- (2a) If you want to meet, I am in my office now.
- (2b) If you will be late, give me a call.
- (2c) If you excuse my saying so, she is downright incompetent.

The form *if A, B* is neither necessary nor sufficient for the expression of conditionality. Inverted forms, as in (3a), are used as conditional antecedents. Sentences like (3b) and (3c) also typically have conditional interpretations.

- (3a) Should the sun come out, Sue will go on a hike.
- (3b) Buy one – get one free.
- (3c) Give me \$10 and I will fix your bike.

On the other hand, some *if-then* sentences do not fit the semantic characterization and are not considered conditionals, as in (4).

- (4) If these problems are difficult, they are also fascinating.

Despite these marginal counterexamples, *if* is clearly the prototypical conditional marker in English. Other languages show more diversity in their expression of conditionality. The German conditional maker *falls* is freely interchangeable with *wenn* ‘when/if’, which also functions as a temporal conjunction. Japanese employs a family of verbal suffixes and particles (*-ba*, *-tara*, *-tewa*, *nara*, *to*), each of which adds subtle semantic and pragmatic constraints to the conditional meaning and some of which may also express temporal relations without conditionality (*-tara* ‘and then’; *A to B* ‘upon A, B’). Languages also vary in the extent to which they overtly mark (non)-counterfactuality. In Japanese, the distinction is usually inferred from context; Classical Greek, on the other hand, has an elaborate inventory of markers of different degrees of hypotheticality.

In all languages, the interpretation of conditionals is determined and constrained by expressions of temporal relations, modality, quantification, and a variety of pragmatic factors. For instance, the differences in (1a) through (1c) arise from the interaction of the marker *if* with the tenses and modal auxiliaries in the constituent clauses.

For descriptive surveys of conditionals in English and other languages, see Traugott *et al.* (1986), Athanasiadou and Dirven (1997), Dancygier (1998), and Declerck and Reed (2001).

Truth-Conditional Semantics

The formal semantic approach in linguistics and philosophical logic is concerned with the truth conditions of sentences and their logical behavior. Conditionals are among the most extensively studied linguistic constructions in this tradition and pose specific challenges, which have been addressed in a number of ways.

Material Conditional

In classical Fregean logic, *if A, B* is interpreted as the *material conditional* (also called material implication) ‘ \rightarrow ’:

- (5) $A \rightarrow B$ is true iff either A is false, or B is true, or both.

The material conditional is a truth function on a par with conjunction and disjunction. However, while there is general agreement that the latter are well suited to capture the truth conditions of *and* and *or*, the logical properties of the material conditional do not well match those of conditional sentences. For example, $A \rightarrow B$ and $A \rightarrow \neg B$ are mutually consistent, and the falsehood of A is sufficient for the truth of both, hence of their conjunction. But (6b) is intuitively contradictory and does not follow from (6a). Likewise, the negation of $A \rightarrow B$ is equivalent to $A \wedge \neg B$, but (6c) and (6d) are not intuitively equivalent.

- (6a) Today is Saturday.
- (6b) If today is Friday, it is raining, and if today is Friday, it is not raining.
- (6c) It is not the case that if the team wins, I will be happy.
- (6d) The team will win and I will be unhappy.

Strictly truth-functional theories employ the material conditional in spite of these shortcomings, since no other truth function comes any closer to capturing our intuitions about conditionals. One way to reconcile the approach with linguistic intuitions is to augment the truth conditions with pragmatic conditions on use. Jackson (1987), building on Grice’s original proposals, appealed to probabilistic ‘assertibility’ conditions. For *if A then B* to be assertible, two conditions must be met: $A \rightarrow B$ must be highly probable, and it must remain highly probable in the event that A turns out true. Jackson noted that this comes down to the requirement that the conditional probability of B given A be high.

(Variably) Strict Implication

An alternative reaction to the problems of the material conditional is to conclude that conditionals do not express truth functions. Instead, most current theories assume that *if A then B* asserts that A *cannot* be true without B also being true. This is typically spelled out in the framework of *possible worlds*:

- (7) *If A then B* is true at a possible world w relative to an accessibility relation R iff for all possible worlds w' such that wRw' and A is true at w' , B is true at w' .

The relation R determines the *modal base* (Kratzer, 1981), the set of possible worlds that are relevant to the truth of the conditional at w . Definition (7) subsumes the material conditional as the special case that R is the identity relation, so the only world

relevant at w is w itself. At the other end of the spectrum lies *strict implication*, under which *all* possible worlds are relevant and the conditional is true iff B is a logical consequence of A .

These extreme cases are rarely relevant in linguistic usage. Usually, conditionals are evaluated against speakers' beliefs, the conversational common ground, the information available in a given situation, possible future courses of events in branching time, or other background assumptions. All of these interpretations correspond formally to different choices of the accessibility relation. The fact that the intended reading need not be overtly marked is a source of versatility and context dependence. A given conditional can be simultaneously true with respect to one modal base and false with respect to another. Thus, (8) may be objectively true, but believed to be false by a speaker with insufficient information or false beliefs.

(8) If this material is heated to 500 °C, it will burn.

The definition in (7) makes room for variation and context dependence of the modal base and overcomes some of the limitations of the material conditional. However, like the latter, it fails to account for the invalidity of certain *nonmonotonic* inference patterns involving conditionals. For instance, under both analyses, a true conditional remains true under Strengthening of the Antecedent (*if A then B* entails *if C and A then B*). Intuitively, however, it is possible for (8) to be true while (9) is false.

(9) If this material is placed in a vacuum chamber and heated to 500 °C, it will burn.

There are several ways of addressing this problem. We will describe two of them, each departing from definition (7) in a different direction.

Relative Likelihood

The first approach takes examples (8) and (9) to show that in cases like (8), not all A -worlds in the modal base are relevant for the truth of the conditional, but only those that satisfy implicit defaults or 'normalcy' assumptions. The listener will assume that air was present (as in [8]) unless this is explicitly denied in the antecedent (as in [9]).

Kratzer (1981) represented such assumptions as an *ordering source*, a set of propositions that are 'normally' true at w . This set induces a preorder on the worlds in the modal base: w'' is at least as normal as w' iff all the propositions in the ordering source that are true at w' are also true at w'' . The interpretation of conditionals is sensitive to the relation in (10).

- (10) *If A then B* is true at w relative to a model base MB iff for every A -world w' in MB , there is an AB -world in MB that is at least as normal as w' and not equalled or outranked in normalcy by any A -world in MB at which B is false.

This offers a solution to the problem posed by (8) and (9). Suppose the material is normally not placed in a vacuum chamber. Then every antecedent-world at which it is, is outranked in normalcy by one at which it is not; thus, (8) may be true while (9) is false.

Formally, the order induced by the ordering source is similar to the relation of 'comparative similarity' between possible worlds that is at the center of the Stalnaker/Lewis theory of counterfactuals (*see* the article **Counterfactuals** for details; Lewis, 1981, for a comparison; and Stalnaker, 1975, for an account of indicative conditionals that refers to this notion). The term 'relative likelihood' is applied to such orders in artificial intelligence (Halpern, 2003). Like the modal base, the ordering source is subject to underspecification and context dependence. Different ordering sources correspond to different readings of the conditional. Besides normalcy, Kratzer (1981) considers ordering sources that rank worlds according to desires, obligations, and other criteria.

Probability

The second approach to dealing with the nonmonotonicity of conditionals does not manipulate the modal base but instead rejects the universal quantification over possible worlds as ill suited for modeling the notion of consequence that speakers employ in interpreting conditionals. On this account, *if A then B* asserts not that all A -worlds are B -worlds but rather that the conditional probability of B , given A , is high. In other words, the posterior probability of B upon learning A would be high, or, alternatively, a world that is randomly chosen from among the A -worlds would likely be one at which B is true. Different modal bases and ordering sources correspond to different (subjective or objective) probability distributions over possible worlds. Adams (1975) developed a theory of *probabilistic entailment* in which just those inference patterns that are problematic for the classical account, such as Strengthening of the Antecedent, are no longer predicted to be valid.

The intuitive appeal of the probabilistic approach is offset somewhat by the fact that it necessitates a rather profound rethinking of the logical basis of semantic theory. Lewis (1976) showed that a conditional probability cannot in general be interpreted as the probability that a proposition is true, hence that the central premise of the probabilistic account is at

odds with the idea that conditionals denote propositions (for detailed discussions see Edgington, 1995; Eells and Skyrms, 1994). Some authors conclude that conditionals do not have truth values (Adams, 1975) or that the conditional probability is only relevant to their use and independent of their truth conditions (Jackson, 1987). Another approach is to assign nonstandard truth values to conditionals in such a way that the problem is avoided (Jeffrey, 1991; Kaufmann, 2005).

Summary

Kratzer's theory is the most influential one in linguistics. The probabilistic approach has been studied extensively in philosophy and, more recently, artificial intelligence. Many other options have been explored. In addition to the works cited above, for overviews and specific proposals the reader is referred to Bennett (2003); Gärdenfors (1988); Harper and Hooker (1976); Harper *et al.* (1981); Jackson (1991); Nute (1980, 1984); Sanford (1989); Stalnaker (1984); Veltman (1985); and Woods (1997). It is not always clear whether there are empirical facts of a purely linguistic nature that would decisively favor one approach over another. With such criteria lacking, the choice depends on the purpose of the analysis at hand and other extralinguistic considerations (e.g., assumptions about rational behavior or psychological reality, or tractability in computational modeling).

See also: Counterfactuals; Formal Semantics; Modal Logic; Possible Worlds; Philosophical Theories.

Bibliography

- Adams E (1975). *The logic of conditionals*. Dordrecht/Boston: D. Reidel.
- Athanasiadou A & Dirven R (eds.) (1997). *Amsterdam Studies in the Theory and History of Linguistic Science, vol. 143: On conditionals again*. Amsterdam/Philadelphia: John Benjamins.
- Bennett J (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.
- Dancygier B (1998). *Conditionals and prediction*. Cambridge: Cambridge University Press.
- Declerck R & Reed S (2001). *Conditionals: a comprehensive empirical analysis*. Number 37 in Topics in English Linguistics. Berlin/New York: Mouton de Gruyter.
- Edgington D (1995). 'On conditionals.' *Mind* 104(414), 235–329.
- Eells E & Skyrms B (eds.) (1994). *Probabilities and conditionals: belief revision and rational decision*. Cambridge: Cambridge University Press.
- Gärdenfors P (1988). *Knowledge in flux: modeling the dynamics of epistemic states*. Cambridge, MA: MIT Press.
- Halpern J Y (2003). *Reasoning about uncertainty*. Cambridge, MA: MIT Press.
- Harper W L & Hooker C A (eds.) (1976). *The University of Western Ontario series in philosophy of science, vol. 1: Foundations of probability theory, statistical inference, and statistical theories of science*. Dordrecht/Boston: D. Reidel.
- Harper W L, Stalnaker R & Pearce G (eds.) (1981). *The University of Western Ontario series in philosophy of science, vol. 15: Ifs: conditionals, belief, decision, chance, and time*. Dordrecht/Boston: D. Reidel.
- Jackson F (1987). *Conditionals*. Oxford/New York: Basil Blackwell.
- Jackson F (ed.) (1991). *Conditionals*. Oxford: Oxford University Press.
- Jeffrey R C (1991). 'Matter-of-fact conditionals.' In *The symposia read at the joint session of the Aristotelian Society and the Mind Association at the University of Durham*. Supp. vol. 65. 161–183.
- Kaufmann S (2005). 'Conditional predictions: a probabilistic account.' To appear in *Linguistics and Philosophy*.
- Kratzer A (1981). 'The notional category of modality.' In Eikmeyer J & Riesner H (eds.) *Words, worlds, and contexts*. Berlin/New York: Walter de Gruyter. 38–74.
- Lewis D (1976). 'Probabilities of conditionals and conditional probabilities.' *Philosophical Review* 85, 297–315.
- Lewis D (1981). 'Ordering semantics and premise semantics for counterfactuals.' *Journal of Philosophical Logic* 10(2), 217–234.
- Nute D (1980). *Topics in conditional logic*. Dordrecht/Boston: D. Reidel.
- Nute D (1984). 'Conditional logic.' In Gabbay D & Guenther F (eds.) *Handbook of philosophical logic, vol. 2: Extensions of classical logic*. D. Reidel. 387–439.
- Sanford D (1989). *If P, then Q: conditionals and the foundations of reasoning*. London/New York: Routledge.
- Stalnaker R (1975). 'Indicative conditionals.' *Philosophia* 5, 269–286.
- Stalnaker R (1984). *Inquiry*. Cambridge, MA: MIT Press/Bradford Books.
- Traugott E C, ter Meulen A, Snitzer Reilly J & Ferguson C A (eds.) (1986). *On conditionals*. Cambridge: Cambridge University Press.
- Veltman F (1985). *Logics for conditionals*. Ph.D. diss., University of Amsterdam.
- Woods M (1997). *Conditionals*. Oxford: Clarendon Press.

Context and Common Ground

H H Clark, Stanford University, Stanford, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

People talking to each other take much for granted. They assume a common language. They assume shared knowledge of such things as cultural facts, news stories, and local geography. If they know each other, they assume shared knowledge of earlier conversations and other joint experiences. And if they are talking face to face, they assume shared knowledge of the scene around them. ‘Common ground’ is the sum of the information that people assume they share. Although the notion is often treated informally, it has a formal definition that has been essential to the study of semantics, pragmatics, and other areas of language.

History

‘Common knowledge’ as a technical notion was introduced by David Lewis (1969) to account for how people coordinate with each other. Suppose A, B, and C agree to meet at city hall at noon. The three of them take it as common knowledge that they intend to go to city hall at noon if and only if: (1) all three believe that the agreement holds; (2) the agreement indicates to all of them that they believe the agreement holds; and (3) the agreement indicates to all of them that they intend to go to city hall at noon. In Lewis’s terminology, the agreement is the ‘basis’ for A, B, and C’s common knowledge that they intend to go to city hall at noon. Common knowledge is always a property of a community of people, even though the community may consist of just two people.

The notion of ‘common ground’ was introduced, in turn, by Robert Stalnaker (1978), based on Lewis’s common knowledge, to account for the way in which information accumulates in conversation:

Roughly speaking, the presuppositions of a speaker are the propositions whose truth he takes for granted as part of the background of the conversation... Presuppositions are what is taken by the speaker to be the **common ground** of the participants in the conversation, what is treated as their **common knowledge** or **mutual knowledge** [p. 320, Stalnaker’s emphases].

In this view, people in conversation take certain propositions to be common ground, and when they make assertions, they add to this common ground. When A tells B, *George arrived home yesterday*, A takes it as common ground with B who George is, what day it is, and where George lives. A uses the

assertion to add to their common ground the proposition that George arrived home the day before. Common ground therefore also includes common (or mutual) beliefs, and common (or mutual) suppositions (Clark and Marshall, 1981; Clark, 1996).

Common ground is a reflexive, or self-referring, notion (Cohen, 1978). If A takes a proposition as common ground with B, then A takes the following statement to be true: A and B have information that the proposition is true **and that this entire statement is true**. (*This sentence has five words* is reflexive in the sense that **this sentence** refers to the sentence that contains it.) Because of the self-reference, people can, technically, draw an infinity of inferences from what they take to be common ground. Suppose A takes it that A and B mutually believe that George is home. A can infer that B believes that George is home, that B believes that A believes that George is home, that B believes that A believes that B believes that George is home, and so on *ad infinitum*. In practice, people never draw more than a few of these inferences. These iterated propositions are therefore a derivative and incomplete representation of common ground. The reflexive notion is more basic (Lewis, 1969; Clark and Marshall, 1981; Clark, 1996).

Bases for Common Ground

In conversation and other joint activities, people have to assess and reassess their common ground, and to do that, they need the right bases. These bases fall into two main categories: community membership and personal experiences (Clark, 1996).

Communal Common Ground

Common ground is information that is common to a community of people. Some of these communities are built around shared practices or expertise, such as the communities of ophthalmologists, New Zealanders, or English speakers. Once A and B mutually establish that they are both ophthalmologists, New Zealanders, or English speakers, they can take as common ground everything that is taken for granted in these communities. Even if A and B mutually establish that A is a New Zealander and B is not, they can take as common ground everything an outsider would think an insider should know about New Zealand. Common ground based on community membership is called ‘communal common ground.’

Everybody belongs to many communities at the same time. Some of these communities are nested

(e.g., North Americans, Americans, Californians, San Franciscans, Nob Hill residents), and others are cross cutting (Californians, lawyers, football fans, Christians). Both nesting and cross-cutting communities lead to gradations in common ground. Any two Californians might readily presuppose common knowledge of the Golden Gate Bridge on San Francisco Bay, but only two San Franciscans would presuppose common knowledge of Crissy Field right next to it.

People have both direct and indirect ways of establishing which communities they jointly belong to. When people meet for the first time, they often begin by exchanging information about their occupations, residences, hobbies, and other identities. They display other communal identities indirectly – in their choice of language, dialect, and vocabulary; their choice of dress and accoutrements; and their age and gender. It is remarkable how many cultural identities people can infer as they talk and how useful these are in establishing communal common ground.

Personal Common Ground

The other main basis for common ground is joint experience. The joint experience may be perceptual. When A and B look at a candle together, they can take their joint experience as a basis for certain mutual beliefs – that there is a candle between them, that it is green, that it smells of bayberry, that it is lit. Or the joint experience may be linguistic or communicative. When A tells B (on April 8), *George arrived home yesterday*, and once they mutually establish that B has understood A, the two of them can take it as common ground that George arrived home on April 7. Common ground that is based on joint perceptual or linguistic experiences between two people is called their ‘personal common ground’. It often holds only for the two of them.

Conversations and other joint activities depend on the orderly accumulation of personal common ground. Suppose A and B are assembling a television stand together. To succeed, they need to establish as common ground what each is going to do next. Part of this they accomplish linguistically, in their spoken exchanges, as when A proposes, *Let's put on this piece next*, and B takes up the proposal, *Okay*. But other parts they accomplish perceptually, as when A hands B a board, screw, or screwdriver, or when A holds up a board and they examine it together. Most face-to-face conversations depend on a mix of linguistic and perceptual bases for the accumulation of personal common ground. Telephone conversations depend almost entirely on linguistic bases.

Language and Communal Common Ground

Communal common ground is fundamental to account for the conventions of language, what are termed the ‘rules of language’. These include conventions of semantics, syntax, morphology, phonology, and pragmatics (Lewis, 1969).

Speakers ordinarily try to use words that their addressees will understand, and that requires a ‘shared lexicon.’ The problem is that every community has its own ‘communal lexicon’ (Clark, 1996). Once A and B jointly establish that they are both speakers of English, they may presuppose common knowledge of a general English-language lexicon. But because other communities are nested and cross cutting, so are the lexicons associated with them. There is a nesting of communities that speak English, North American English, New England English, and Bostonian. Although words such as *dog* and *in* are common to English in general, others are common only to one or another nested community; in Bostonian, for example, a *barnie* is a Harvard student. Indeed, every community (Californians, lawyers, football fans, ophthalmologists) has a specialized lexicon. The lexicon for lawyers includes *tort*, *mortmain*, and *ne exeat*. The lexicon for ophthalmologists includes *tonometry*, *uvea*, and *amblyopia*. To use *barnie* or *mortmain* is to take as common ground a Bostonian or legal lexicon. Communal lexicons are sometimes called jargon, dialect, patois, idiom, parlance, nomenclature, slang, argot, lingo, cant, or vernacular; or they consist of regionalisms, colloquialisms, localisms, or technical terminology.

Speakers also try to use syntactic constructions, or rules, that they share with their addressees. For example, in English generally, it is conventional to mention place before time (*George is going to London tomorrow*); yet in Dutch, a closely related language, it is conventional to mention place and time in the reverse order (*Pim gaat morgen naar London*, ‘Pim goes tomorrow to London’). The rules of syntax, however, vary by nested communities. It is conventional to say *He gave it me* in British English, but not in English generally. It is conventional to say *My car needs washed* in Western Pennsylvania English, but not in North American English. Many rules of syntax are tied to specific words in a communal lexicon, and these vary from one community to the next.

Speakers also try to use, or adapt to, the phonology of their cultural communities. Indeed, pronunciations vary enormously from one community to the next. The vowel in *can't*, for example, changes as one goes from British to North American English, from northern to southern dialects of American English, and

even from one social group to another within a single school. Also, the same person may pronounce *singing* as 'singing' in an informal setting but as 'singing' in a classroom or a court of law.

Discourse and Personal Common Ground

Personal common ground is essential to the processes by which people converse. To *communicate* is, according to its Latin roots, to *make common* – to establish something as common ground. To succeed in conversation, people must design what they say (1) against the common ground they believe they already share with their interlocutors and (2) as a way of adding to that common ground (Stalnaker, 1978). Two consequences of trying to make something common are 'information structure' and 'grounding.' 'Information structure' is a property of utterances. When A tells B, *What the committee is after is somebody at the White House*, A uses the special construction to distinguish two types of information (Prince, 1978). With the Wh-cleft *What the committee is after*, A provides information that A assumes B is already thinking about. It is one type of 'given information.' In contrast, with the remainder of the utterance 'is somebody at the White House,' A provides information that A assumes B doesn't yet know. It is 'new information.' Given information is assumed to be inferable from A and B's current common ground, whereas new information is not. New information is, instead, what is to be added to common ground. The way people refer to an object in a discourse (e.g., the committee, somebody, of the White House) depends on whether they believe that the object is readily evoked, known but unused, inferable, or brand new in their common ground for that discourse (Prince, 1981).

'Grounding' is the process of trying to establish what is said as common ground (Clark and Schaefer, 1989; Clark and Brennan, 1991). When A speaks to B in conversation, it is ordinarily not enough for A simply to produce an utterance for B. The two of them try to establish as common ground that B has understood what A meant by it well enough for current purposes. In this process, B is expected to give A periodic evidence of the state of his or her understanding, and A is expected to look for and evaluate that evidence. One way B can signal understanding is with back-channel signals such as *uh-huh*, *yeah*, a head nod, or a smile. Another way is with the appropriate next contribution, as when B answers a question asked by A. But if B does not manage to attend

to, hear, or understand A's utterance completely, the two of them will try to repair the problem. One way is illustrated here:

A (on telephone): Can I speak to Jim Johnstone, please?

B: Senior?

A: Yes.

B: Yes.

In turn 2, B asks A to clear up an ambiguous reference in A's question, and in turn 3, A does just that. Only then does B go on to answer A's question. Turns 2 and 3 are called a 'side sequence' (Jefferson, 1972). Grounding takes many other forms as well.

Common ground is central to accounts of language and language use. It is needed in accounting for the conventions, or rules, of language and to explain how people contribute to conversation and to other forms of discourse.

See also: Contextualism in Epistemology; Conventions in Language; Presupposition.

Bibliography

- Clark H H (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark H H & Brennan S A (1991). 'Grounding in communication.' In Resnick L B, Levine J M & Teasley S D (eds.) *Perspective on socially shared cognition*. Washington, DC: APA Books. 127–149.
- Clark H H & Marshall C R (1981). 'Definite reference and mutual knowledge.' In Joshi A K, Webber B L & Sag I A (eds.) *Elements of discourse understanding*. Cambridge: Cambridge University Press. 10–63.
- Clark H H & Schaefer E R (1989). 'Contributing to discourse.' *Cognitive Science* 13, 259–294.
- Cohen P R (1978). On knowing what to say: planning speech acts. Ph.D. diss., University of Toronto.
- Jefferson G (1972). 'Side sequences.' In Sudnow D (ed.) *Studies in social interaction*. New York: Free Press. 294–338.
- Lewis D K (1969). *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.
- Prince E F (1978). 'A comparison of Wh-clefts and It-clefts in discourse.' *Language* 54(4), 883–906.
- Prince E F (1981). 'Towards a taxonomy of given-new information.' In Cole P (ed.) *Radical pragmatics*. New York: Academic Press. 223–256.
- Stalnaker R C (1978). 'Assertion.' In Cole P (ed.) *Syntax and semantics 9: Pragmatics*. New York: Academic Press. 315–332.

Context Principle

R J Stainton, University of Western Ontario, London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

It is a near truism of the philosophy of language that a word has meaning only in the context of a sentence; this principle is sometimes formulated as the claim that only sentences have meaning in isolation. This is the context principle, first emphasized in Western philosophy by Frege (1884), endorsed early on by Wittgenstein (1922: 51), and sanctioned more recently by Quine (1951: 42), among many others. The Principle and several different ways of understanding it seem to have been foreshadowed in classical Indian philosophy. (See also Matilal and Sen, 1988.)

In this article, I provide some background to the Principle, describe three ways of reading it (a methodological reading, a metasemantic reading, and an interpretational/psychological reading). I offer some reasons for endorsing the Principle, and some reasons for being skeptical.

The heated exegetical controversies over Frege's relationship to the Principle are not presented in this article. Some believe that Frege would have applied it to both sense and reference; others disagree. Some believe that Frege rejected the Principle in his later work, others that he retained it throughout. In addition, different authors take Frege to endorse different readings of the Principle: nearly everyone would agree that he accepted the methodological reading, but it is less clear whether he endorsed the metasemantic or interpretational/psychological reading. Such scholarly issues are not my concern in this article. For a thorough discussion, see Dummett (1981: 369ff, 1993a).

Sentence Primacy: Three Interpretations of the Context Principle

The context principle gives primacy to sentences. Specifically, sentences are taken to be semantically prior to the words that make them up. The Principle is, in this regard, a member of a family of theses that have some whole to being somehow 'prior' to its parts. As with all such doctrines, one obtains a holistic primacy thesis by specifying what the whole is, what its parts are, and in what sense the former is prior to the latter. Most important for present purposes, one can mean different things by 'prior.' Of particular interest here, one can take sentences to be methodologically prior, metasemantically prior, or interpretationally prior to the words that compose them.

Let me begin with the methodological reading of the Principle. In his *Foundations of Arithmetic*, Frege (1884: x) famously promised to keep to the following fundamental constraint: "never to ask for the meaning of a word in isolation, but only in the context of a sentence." Taken as a methodological precept, this principle essentially tells the lexical semanticist only to contemplate the effect that a word can have on sentences in which it may be embedded. For instance, to find out the meaning of the word 'one' (an example of great interest to Frege), the lexical semanticist should reflect upon such questions as the following: What whole sentences containing 'one' have in common (e.g., "One apple fell" and "One dog died"); how sentences that contain words slightly different from 'one' differ systematically in meaning from maximally similar sentences containing 'one' (e.g., "One dog died" versus "No dog died"); and so on. What the lexical semanticist should never do is try to figure out the meaning of 'one' just by thinking about it – that phrase – in isolation (where in isolation means not embedded in any larger syntactic structure).

The second reading of the context principle considered in this article is the metasemantic reading. A metasemantic view is a view about the source of meaning. It poses an "in virtue of what" question. Here's an example. Suppose we ask

- (1) In virtue of what is the sound /to:fu/ meaningful?
In virtue of what does it mean "a pale curd of varying consistency made from soybean milk," rather than "sea lion" or "watch"?

Notice that we are not asking, in (1), what the sound /to:fu/ means. Rather, we are asking why it means what it does. Nor is this the causal-historical question about the steps whereby /to:fu/ came to have this meaning. It is, instead, the issue of what more primitive present facts make for this less primitive present fact: how do the 'higher' facts emerge from 'lower' ones? For example, compare these two questions: what makes it the case that things have the monetary value they do, or what makes it the case that certain things are illegal, or rude, or immoral? These too are "in virtue of what" questions.

Some philosophers seem to have taken from Frege's discussion of "not asking for the meaning of a word in isolation" a claim about what makes words meaningful and what makes them have the meaning they do. The claim is that, fundamentally, only sentences have meaning. This is not to say that subsentences are gibberish. Rather, the entities that have meaning in the first instance are sentences. Unlike the first reading of the Principle, this doctrine is not about where one should look to find out about meaning;

it is, rather, a doctrine about where meaning comes from, i.e., the basic source of meaning. What the Principle says is that the only things that have meaning non-derivatively are sentences, so it must be in virtue of their role within sentences that subsentential expressions have meaning at all.

Here is the same idea put another way: suppose that some expressions obtain their meaning from how they alter the meanings of larger wholes. Suppose, indeed, that this is how words/phrases obtain their meaning; they therefore have meaning only derivatively, not fundamentally. Now, it cannot be the case that all expressions obtain their meaning in this way or there would be an infinite regress. The claim says that the things that have meaning non-derivatively are sentences.

Does this mean that one must first grasp the meaning of each of the infinite number of sentences in the language and only then solve for word meanings? No, not least because doing so is not humanly possible. To avoid this problem, proponents of the metasemantic version of the context principle can make several claims. First, they may insist on a sharp difference between (1) a psychological story about how humans grasp word and sentence meanings and (2) a philosophical story about the metaphysical underpinnings of word and sentence meaning. They may then eschew any claims about the first of these, stressing that they only mean to address the second (see Dummett, 1973: 4 for this approach). Second, the proponents of the context principle, read metasemantically, could propose that there is some finite cluster of simple sentences, the meaning of which one grasps from use; one then presumably solves for the meaning of the words and for the contribution of syntax, using just those sentences. Performing this finite task then gives the person the capacity to understand new sentences, a potential infinity in fact, on the basis of the (familiar) words in the (unfamiliar) sentences and how those words are structured. Either move would save the proponents of the metasemantic thesis from endorsing the absurd view that one first understands all sentences and only then understands any words.

So far we have examined two readings of the context principle. The first was merely methodological, a claim about how to find out what particular words mean: To find word meanings, look at what they contribute to sentences. The second reading was metasemantic, a claim about why words have the meanings they do: words only have meaning because of how they affect sentence meanings. The third reading of the Principle is interpretational/psychological. It is an empirical claim about the psychology underlying comprehension. Dummett (1993b: 97) discusses

the view that “it is possible to grasp the sense of a word only as it occurs in some particular sentence.”

In a way, this reading of the Principle is the most straightforward of the three: the idea underlying it is that the only things we are psychologically able to understand are whole sentences. Put in terms of generative capacity, the claim would amount to this: the only thing that our semantic competence generates are meanings for whole sentences; it does not output meanings for words/phrases (though it presumably uses word/phrase meanings in generating meanings for whole sentences, they just are never ‘output’). Thus, we can understand words only when they are spoken within whole sentences. Even this most straightforward of the three readings admits of further subreadings, however. Dummett (1993b: 109), for instance, contrasted two varieties of “grasping a sense,” one dispositional and the other occurrent. He granted that one may dispositionally grasp the sense of a subsentence outside the context of any sentence. However, he apparently denied – or anyway, has Frege deny – that one can, in the occurrent sense, grasp the sense of a word/phrase without grasping the sense of a sentence within which that word/phrase occurs. This would mean that one could “know the meaning” of a word in isolation, but that whenever one put that knowledge to work, in actual understanding, it would have to be in grasping a sentential content. This last is what the context principle would come to, on this weaker subreading of the interpretational/psychological principle.

Motivating the Context Principle

Having explained three senses in which one could take whole sentences to be prior to the words that make them up, let us consider reasons for endorsing sentence primacy. Some of these reasons support just one reading of ‘priority.’ Some support more than one. Given the limited space, I present only three such reasons and for the most part leave for the reader the question of which reason supports which reading of “Sentences are prior.”

Frege believed that, in failing to obey his methodological constraint, “one is almost forced to take as the meanings of words mental pictures or acts of the individual mind” (Frege, 1884: x). Thus in the case of number-words, the failure to respect the principle could easily lead one to suppose that ‘one’ stands for a mental item, and hence that mathematics is somehow about mental entities, which in Frege’s view is an extremely serious error (see Frege, 1884: 116). However, when one obeys the principle, one comes to the right view: the meaning of a word is not some idea that we associate with it, but is instead the thing

that the word contributes to the meaning of larger expressions. Frege writes (1884: 71):

That we can form no idea of its content is therefore no reason for denying all meaning to a word, or for excluding it from our vocabulary. We are indeed only imposed on by the opposite view because we will, when asking for the meaning of a word, consider it in isolation, which leads us to accept an idea as the meaning. Accordingly, any word for which we can find no corresponding mental picture appears to have no content. But we ought always to keep before our eyes a complete proposition. Only in a proposition [Satz] have the words really a meaning.

So, one advantage of endorsing the Principle is that it keeps us from making such a mistake.

Consider this related motivation: starting from the top – focusing on whole sentence meanings and only then considering what the parts must mean, in order for the observed whole meaning to be generated – opens up the possibility of novel and surprising accounts of what the parts mean. Indeed, it becomes possible to conceive of syntactic parts that, though they have some sort of impact on meaning, do not themselves have a meaning in isolation. Such parts receive only what is called a ‘contextual definition.’ This concept is best explained by appeal to an example. If we start by looking at the phrasal parts of “The king of France is bald,” asking what they mean, it can seem inevitable that the phrase, “The king of France,” must stand for an object. What else could its meaning be, in isolation? This, of course, raises all manner of ontological issues: What is this bizarre object, since there is, in reality, no king of France? How can such an unreal entity be bald or not, so as to render this sentence true or false? And so on. Crucially, however, if we pursue the methodology suggested here and start with the whole sentence, we may notice, with Russell (1905), that the sentence as a whole means the following: there is exactly one king of France, and every king of France is bald. We may further notice that this whole meaning can be generated without assigning any reference at all to the phrase, “The king of France.” This is not to say that this phrase makes no difference to what the whole means; patently it does make a difference. However, in place of a meaning-entity for “The king of France,” all we need is a rule, a contextual definition, that says:

- (2) A sentence of the form “The *F* is *G*” is true iff exactly one thing is *F* and everything that is *F* is *G*.

Taking this contextual definition to be the meaning-determining rule, we simply avoid the issue of what the phrase, “The king of France,” stands for, since the phrase itself, upon analysis, does not contribute a constituent to the whole meaning.

Another methodological advantage of the context principle, then, is that it is rather easier to arrive at this kind of contextual definition than if we begin with what the parts mean, in isolation.

A second kind of advantage is that, by strictly obeying the context principle, we automatically meet a key constraint of semantic theories: compositionality. Roughly speaking, compositionality says that the meaning of a whole expression is exhausted by (1) what its parts mean, and (2) how those parts are put together (*see Compositionality: Semantic Aspects, Compositionality: Philosophical Aspects* for more details.) Compositionality is accepted as a constraint for two related reasons. First, insofar as these are the sole determinants of whole meanings, we can explain why people understand complex expressions that they have never encountered before: they understand them by calculating the whole meaning from precisely these two elements, both of which are familiar. Second, were whole meanings not compositional, it would be an utter mystery how we finite beings could in principle know the meaning of the infinite number of sentences that, though we have never heard them, we would, but for our finite lifetime and memory, be capable of understanding. That is, compositionality accounts for an observed ability in practice and a different though related ability in principle. Notice, however, that compositionality is one side of a coin, the other side of which is the context principle. Compositionality says that whole meaning is entirely a function of part meanings plus structure:

- (3) Whole meaning = $\langle \text{part-meaning}_1, \text{part-meaning}_2, \dots, \text{part-meaning}_i, \dots, \text{part-meaning}_n \rangle + \text{structure}$

The context principle employs this same equation to solve for a part meaning, i.e., taking part meaning to be entirely determined by the whole meaning, the meanings of the other parts, and the structure:

- (4) $\text{Part-meaning}_i = \text{Whole meaning} - \langle \text{part-meaning}_1, \text{part-meaning}_2, \dots, \text{part-meaning}_n \rangle + \text{structure}$

So, if we assign part meanings in line with (4), the context principle, we cannot help but get the desired result vis-à-vis (3), i.e., compositionality. (Note: obviously the manner of combination of part meanings and structure is not literally addition. Nevertheless, I use the symbols ‘+’ and ‘−’ to simplify presentation.) Automatically satisfying the compositionality constraint in this way is thus another advantage of endorsing the context principle.

A third kind of motivation for endorsing the Principle is that it seems to be connected with several

other holistic primacy theses, each of which is allegedly independently motivated. (Unfortunately, space does not permit me to explain what the independent motivation is for these other theses. See Brandom, 1994, chapter 2, sections II and III, for discussion and an overview of the relations among these various primacy claims.) Kant (1787) famously insisted that judgment is prior to perception of individuals: seeing that María is a female, a person, tall, and the like is prior to seeing María. Put otherwise, whereas classical empiricists started with representations of individual objects and of universals and then built up complex mental representations that could be true/false, Kant turned this on its head: the whole representation (i.e., what is judged) is prior to the object-denoting parts that make it up. The early Wittgenstein (1922: 31) also insisted that facts are prior to the objects and properties that make them up: “the world is the totality of facts, not of things.” In a related move, Dummett (1973) has urged, following the later Wittgenstein, that the practice of assertion – and other full-fledged “moves in the language game” – is prior to the act of referring. As Wittgenstein (1953: 24) put it:

For naming and describing do not stand on the same level: naming is a preparation for description. Naming is so far not a move in the language-game – any more than putting a piece in its place on the board is a move in chess. We may say: *nothing* has so far been done, when a thing has been named. It has not even *got* a name except in the language-game. This was what Frege meant too, when he said that a word had meaning only as part of a sentence.

Adopting these primacy theses can, each in their own way, lead one to expect sentences to be primary as well. Goes the idea, what is judged are sentential representations; the linguistic item that corresponds to a fact is a sentence, and the linguistic item that we assert with is the sentence.

Dummett’s point about sentence use deserves to be expanded upon, since it underlies several of the points made above. Dummett suggested that the only things that can be used in isolation – that is, used without being embedded in a larger structure – are sentences. He wrote (1973: 194):

A sentence is, as we have said, the smallest unit of language with which a linguistic act can be accomplished, with which a “move can be made in the language-game”: so you cannot *do* anything with a word – cannot effect any conventional (linguistic) act by uttering it – save by uttering some sentence containing that word.

Yet, as a famous Wittgensteinian slogan says, meaning comes from use (see Wittgenstein, 1953 and elsewhere). Thus, the things that have meaning

fundamentally have it because of their use: an expression has the non-derivative meaning that it does because of the kinds of actions speakers can perform with it. However, as suggested just above, those just are the sentences. So words must get their meaning because they appear in meaningful sentences. Dummett, expanding on this Wittgensteinian theme, put the general lesson as follows:

Indeed, it is certainly part of the content of the dictum [i.e., the context principle] that sentences play a special role in language: that, since it is by means of them alone that anything can be *said*, that is, any linguistic act (of assertion, question, command, etc.) can be performed, the sense of any expression less than a complete sentence must consist only in the contribution it makes to determining the content of a sentence in which it may occur (1973: 495; see also Dummett, 1993a).

A Possible Objection to the Context Principle

Having noted three kinds of reasons for embracing the context principle, let me end with an objection that may come immediately to mind. First, it seems that adults speak in subsentences all the time. I see a woman wearing a lovely garment and say to my wife, “Nice dress.” I receive a letter in the mail, hold it up, and say to my companion, “From Spain.” Such talk is absolutely ubiquitous. (For empirical support, see the papers in Elugardo and Stainton, 2004, and the many references cited there; for an overview, see Stainton, 2004.) Second, children learning a language seem to start with subsentences – which makes it equally hard to see how grasping a sentential meaning could be a prerequisite for grasping a subsentential one. Let us consider the problem that such subsentential speech might pose for the Principle.

Start with the methodological reading. It is a bit strong to demand that one never consider the word in isolation if words/phrases can be used unembedded to perform speech acts. More appropriate, and still in the broadly Fregean spirit, would be this claim: never *only* consider the word in isolation, but instead *also* consider its behavior when embedded in whole sentences. Non-sentential speech does not conflict with this latter, more inclusive, methodological precept. In addition, the methodological point of the context principle – to cure one of the habit of taking mental images and such as meanings – is met even on this weaker reading. Hence subsentence use actually poses no problems for the Principle, on this first reading.

What of the metasemantic doctrine? Notice that a key premise in the argument for the doctrine was

that only sentences can be used to perform speech acts. Words and phrases cannot be: that is why they were denied meaning, fundamentally speaking. Yet, this key premise looks false, if words really can be used in isolation. Therefore, without this premise, some other argument must be given for the conclusion that only sentences have meaning fundamentally. Thus subsentence use, if genuine, does not falsify the Principle read in this way, but it does leave one in need of an empirically adequate argument for meaning having to come from sentences alone.

It might seem that a better argument for the claim that meaning must still come from sentences is at hand: Surely this doctrine is required to preserve compositionality. As I stressed above, you do not get (3) above unless you also accept (4) and (4) requires that word meanings – the meaning of the parts – not exceed what they contribute to full sentences. In fact, however, compositionality does not, on its own, support the metasemantic doctrine, which makes two claims: first, sentences are a metaphysical *source* of word meaning, and second, they are the *only* such source. Neither of these claims, however, can be inferred from compositionality *per se*. All (4) gives us is a constraint: Whatever story we tell about where a word's meaning comes from, it must be consistent with sentence meanings being exhausted by what their parts mean. This does not support any claim about sources. Moreover, if words are used in isolation, then, though sentence use might be one source, it surely would not be the only one.

To see why compositionality does not, taken alone, support the metasemantic doctrine, consider an analogy. Take this proposal: facts about what art works are beautiful derive from facts about what works are attractive to (most) art experts. That is, it is in virtue of the judgment of (most) experts that art works are beautiful or not. Suppose one tried to defend this meta-esthetic view by saying: "Look, it can't be that most genuine experts are wrong about what's beautiful. They wouldn't be experts otherwise." This defense would not really succeed as an argument for the meta-esthetic view because, even granting it, one could only infer that it is a constraint on where beauty comes from that most experts are right about what is beautiful. This fact would not, on its own, support the idea that beauty comes from expert judgment. Nor would it support the even stronger idea that beauty comes solely from expert judgment. In the same way, compositionality may well impose a constraint on metasemantic theories: one might well contend that any successful metasemantics must have whole meanings exhaustively determined by part meanings and linguistic structure. Yet, one cannot go from such a constraint

immediately to conclusions about where meaning-facts emerge from; still less can one move from such a constraint to a conclusion about the sole thing from which they emerge. In sum, given subsentential speech, we are still in need of a reason for embracing the metasemantic reading of the context principle.

Let me now make a brief detour into a related issue. One reason that it matters whether the metasemantic doctrine is upheld is this: If sentence meaning is the only source of word meaning, then it is arguable that the latter is indeterminate. That is, there might be no fact of the matter about what individual words "really mean." The argument goes like this. We can hold constant the meaning of every sentence in the language while varying the contribution that we assign to the words within those sentences. To give a highly simplified example, one way to assign the right meaning to the Spanish sentence, "María no fuma" ["María doesn't smoke"] is to assign the person MARIA to 'María', SMOKEs to 'fuma', and DOESN'T to 'no'. Another way, which still gives the right meaning for the whole sentence, is to assign the person MARIA to 'María no' and DOESN'T SMOKE to 'fuma'. Now, with respect to this highly simplified example, we can find reasons for picking the first over the second option: 'fuma', 'no' and 'María' show up in lots of sentences, and their contribution in those other sentences is, surely, SMOKEs, DOESN'T, and MARIA, respectively. So that is what they contribute here too. However, suppose we revised our view of the meaning of the other parts in all sentences containing 'fuma', 'María' and 'no'. Surprisingly, it has been suggested that this sort of rearrangement is something we could systematically do. The result would be that the complete set of sentences containing a given word leaves us with various options about what the word means. Further, assuming that the meaning of all sentences in which a word occurs is the sole thing that metaphysically determines its meaning, there can be no single thing that is "the meaning of 'fuma'." This is the thesis of indeterminacy (see Quine, 1960 and Putnam, 1981 for worked-out examples).

I introduce the indeterminacy thesis because it highlights the sense in which the metasemantic version of the context principle says more than "the meanings one assigns to words must fit with the meanings one assigns to sentences containing those words." It also says that the word meanings are *exhausted* by sentence meanings – in a way that can lead to indeterminacy. In contrast, if word meanings depend also upon how words are used on their own, then even if the complete set of sentence meanings does not fix the meaning of individual words, we cannot yet conclude that word meaning

is indeterminate. For word meaning might be more completely fixed by how words in isolation are used (for more on this connection between the context principle and indeterminacy, see Stainton, 2000).

We have seen that subsentence use is consistent with the methodological reading of the context principle. It is also consistent with the metasemantic reading, though it leaves this latter doctrine in need of an empirically adequate supporting argument. Consider finally the interpretational/psychological doctrine. It says that, as a matter of our psychology, we cannot understand a word, when uttered, unless it is embedded in a sentence. This reading of the context principle seems simply false, given the existence of subsentential speech. There is no hope for making it consistent with genuine subsentence use. Apparently, hearers understand subsentential expressions in isolation; hence their semantic competence must generate a meaning for such expressions in isolation. The best hope for the Principle read in this strongest way is thus to deny that the phenomenon of subsentential speech is genuine: adults do not actually speak in subsentences, they merely appear to do so. What is really going on is that adults speak ‘elliptically’ in some sense – they produce sentences, but those sentences somehow “sound abbreviated” (see Stanley, 2000 for this sort of idea). As for children, who seem to grasp word meanings long before they grasp the meanings of any sentences, proponents of the interpretational reading of the context principle must make some fairly implausible suggestions. They may insist that children actually do understand sentence meanings even though they do not speak in sentences; or they may claim that what children mean by their words (e.g., ‘doggie’) is not what the adult word means. The child’s expression, they might insist, is actually a one-word sentence meaning, “There is a dog,” and hence is not synonymous with our word. (That is, on this second disjunct, the idea would be that children actually do not employ/understand *our* words outside sentences, but rather they employ homophonous sentences – until, that is, they are also competent with *our* sentences.)

Does this inconsistency with the interpretational/psychological reading mean that the other primacy doctrines – of judgment, facts, and assertion – are also required to make these implausible empirical claims? After all, it was suggested that those doctrines supported sentence primacy. The answer is no, because these other primacy doctrines really do not entail anything about only sentences being used and only sentence meanings being graspable occurrently. At best what they lend credence to is the primacy of a certain sort of content, namely the proposition. For, strictly speaking, it is propositions that are judged,

propositions that correspond to facts, and propositions that are exchanged in assertion. Further, subsentential speech does not call the centrality of propositions into question: When I say “Nice dress” or “From Spain,” I still convey something propositional; that is, a proposition about the salient dress to the effect that it is nice, and a proposition about the letter to the effect that it is from Spain, respectively. I merely do so using *linguistic expressions* that are not propositional. So, subsentential speech leaves proposition primacy intact. To move immediately and without further argument to any conclusion about the syntactic structures that (purportedly) express propositions, however, is to commit some kind of global use/mention error, running together features of a content (i.e., a proposition) with features of its supposed linguistic ‘vehicle’ (i.e., a sentence). In short, even if one takes judgments, facts, or assertions to be primary, one need not endorse the context principle vis-à-vis interpretation – since the latter is about the centrality of a certain class of syntactic items.

In summary, I have presented three different ways of reading the context principle: methodological, metasemantic, and interpretational/psychological. I then noted three rationales for embracing the Principle: to avoid the errors of psychologism, to enforce compositionality, and because of links to other independently motivated ‘primacy doctrines.’ I ended with an objection to the Principle, from non-sentence use. The suggested result, in the face of this objection, was two parts consistency and one part inconsistency: (1) the first reading of the Principle would be untouched, (2) the second would be left unsupported, but (3) the third reading would be outright falsified, so that the proponent of this reading of the Principle must make some (implausible) empirical claims to the effect that people do not actually speak subsententially.

See also: Compositionality: Philosophical Aspects; Compositionality: Semantic Aspects; Holism, Semantic and Epistemic; Indeterminacy, Semantic.

Bibliography

- Brandom R (1994). *Making it explicit*. Cambridge, MA: Harvard University Press.
- Dummett M (1973). *Frege: philosophy of language*. Cambridge, MA: Harvard University Press.
- Dummett M (1981). *The interpretation of Frege’s philosophy*. London: Duckworth.
- Dummett M (1993a). ‘The context principle: centre of Frege’s philosophy.’ In Ingolf M & Stelzner W (eds.) *Logik und mathematik*. Berlin: De Gruyter. 3–19.

- Dummett M (1993b). *Origins of analytical philosophy*. Cambridge, MA: Harvard University Press.
- Elugardo R & Stainton R (eds.) (2004). *Ellipsis and non-sentential speech*. Dordrecht: Kluwer.
- Frege G (1884). *Foundations of arithmetic* (2nd rev. edn., 1978). Austin J L (trans.). Oxford: Blackwell.
- Kant I (1787). *Critique of pure reason*. Smith N K (trans., 1929). New York: St. Martin's Press.
- Matilal B K & Sen P K (1988). 'The context principle and some Indian controversies over meaning,' *Mind* 97, 73–97.
- Putnam H (1981). *Reason, truth and history*. Cambridge: Cambridge University Press.
- Quine W V O (1951). 'Two dogmas of empiricism.' *From a logical point of view*. Cambridge, MA: Harvard University Press. 20–46.
- Quine W V O (1960). *Word and object*. Cambridge, MA: MIT Press.
- Russell B (1905). 'On denoting,' *Mind* 14, 479–494. Reprinted in Marsh R C (ed.) (1956). *Logic & Knowledge*. London: Unwin Hyman.
- Stainton R J (2004). 'The pragmatics of nonsentences.' In Horn L & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 266–287.
- Stainton R J (2000). 'The meaning of "sentences."' *Nous* 34(3), 441–454.
- Stanley J (2000). 'Context and logical form.' *Linguistics & Philosophy* 23, 391–434.
- Wittgenstein L (1922). *Tractatus logico-philosophicus*, tr. Ogden C K. London: Routledge & Kegan Paul.
- Wittgenstein L (1953). *Philosophical investigations*. Oxford: Blackwell.

Contextualism in Epistemology

H Jackman, York University, Toronto, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

For a subject to truly be said to 'know' a proposition p , p has to be true, and the subject has to have a belief that p meets certain standards of justification. The invariantist will argue that, for any given proposition, the *same* standards are always in place. (The skeptical invariantist will usually take these to be very high standards, while the dogmatist will usually take them to be comparatively low.) By contrast, contextualists argue that the standards for the truth of knowledge attributions can vary from context to context. Consequently, in some contexts being able to answer skeptical doubts about our being, say, brains in vats tricked into thinking that we are experiencing a physical environment rather than a machine simulation, may be required to count as knowing that we have hands, while no such requirement may be in place in other contexts.

According to the contextualist, just as context shifts the satisfaction conditions of words like 'he' or 'tall,' context can shift the satisfaction conditions for 'know.' Contextualism has considerable appeal, since it seems clear that the standards associated with our use of 'know' show *some* sort of contextual variation. In everyday contexts, we claim to know all sorts of things; if the stakes are higher, we may claim to know less, and when faced with skeptical arguments, we may admit to knowing almost nothing. Contextualists can take the truth conditions of knowledge claims to vary the way our practice suggests. It thus both explains the intuitive pull of skepticism (skeptical arguments raise the standards for

knowledge attribution) and allows that the soundness of skeptical arguments doesn't undermine the truth of everyday knowledge claims.

Nevertheless, invariantists have argued that the contextual variation in our use of 'know' should not be reflected in the term's semantics. Instead, they suggest that some aspect of our practice (either the everyday knowledge claims or our capitulation to skeptical arguments) is either (1) systematically in error, or (2) motivated more by pragmatic rather than semantic considerations. Both strategies have the advantage of positing a simpler, context-invariant, semantics for 'know,' but the required supplementary explanations of why our use of the term varies from the proposed semantics have often seemed suspiciously *ad hoc*.

However, even if we were to admit that the semantics of 'know' is context sensitive in some way, there is still a question of which context it is sensitive to. Three possible candidates for this context are the following:

1. The subject's context: The context in which the purported knower is in.
2. The attributor's context: The context in which the claim is made that the subject does, or does not, know a given proposition.
3. The assessor's context: The context in which the claim about the subject's (lack of) knowledge is evaluated as either true or false.

When one considers one's own first-person present-tense knowledge claims such as "I know that it's raining," all three contexts can overlap. However, if A claims, "It's raining," and B later claims, "A knew that it was raining," the subject's and attributor's

context will be different. Furthermore, if C goes on to claim that “B’s claim that A knew that it was raining was false,” the context of assessment will be different from the former two as well.

How radical a doctrine contextualism will turn out to be will depend largely on which context knowledge claims are taken to be relative to.

The least radical of these takes the subject’s context to be the relevant one. For instance, if A sees his cell phone across the table, and nothing much hangs on the issue, then he may count as knowing that his phone is nearby. On the other hand, if A is expecting a particularly urgent call on his phone, he might not count as knowing without picking the phone up to make sure that it is his rather than just another phone of the same type. Still, there will be nothing subjective about whether or not a person counts as knowing a given proposition. If the subject is in the low-standards context, then anyone who denies that the subject knows that his phone is in the room will be mistaken, no matter what context the particular attributor is in.

However, subject-based contextualism does not seem to capture the way in which the standards associated with our use of ‘know’ seem to vary. For instance, if the subject’s context changes such that he will no longer be able to claim that he knows that his phone is across the room (say, he receives an e-mail telling him to expect the urgent call), he will not say that, five minutes prior to this change he did know, but now he no longer does. Rather, he will now say that he didn’t know before either, which suggests that the standards vary not according to the context that the subject is in, but rather with something else (DeRose, 2002; Hawthorn, 2004).

Of course, it will be possible to explain the differences between these aspects of our use of ‘know’ and the first sort of contextualism by appealing to linguistic error or pragmatic factors to explain the divergence, but to the extent to which one claims that we are simply mistaken about the truth conditions of our knowledge claims, the contextualist’s methodological advantage over the invariantist will seem to disappear.

On the other hand, cases like that just mentioned can be handled if we take the context of attribution to be the relevant one. On such attributer-based contextualism, it will be possible for two attributors to make seemingly conflicting true claims about whether A knows that his cell phone is in the room. If attributor B is in a low-standards context, he may truly claim that A knows that his phone is in the room, while if attributor C is in a context with higher standards, he may truly claim that A does not know this. The standards that are relevant for knowledge

are determined not (exclusively) by A’s situation, but also by B’s and C’s contexts, and considerations that are relevant for B and C might not be so for A. Knowledge attributions will thus be in an important sense, relative, since B and C can make seemingly conflicting true claims about A. The relativism involved here is, however, comparatively harmless, since there is a good sense in which B’s and C’s claims do not really conflict. Again, ‘know,’ on such an account, would be context sensitive in the way that, say, ‘tall’ is. If my 10-year-old cousin is 5’6”, he may be described as “tall” when we are discussing the prospects of his third-grade basketball team, or he may be described as “not tall at all” when we are discussing which members of my family could still sit comfortably in the back seat of a compact car. The second claim needn’t contradict the first, and many would say that both claims can be true, it’s just that the standard for being ‘tall’ is different in the two contexts.

While such attributer-based contextualism is probably the most popular version of the view (DeRose, 1992, 1995, 2002; Lewis, 1996; Neta, 2002, 2003; Williams, 1991), it fails to track some fairly well-entrenched aspects of our use of ‘know.’ In particular, when the standards for knowledge attributions go up, not only do we (1) refuse to claim that we know things that we claimed to know before, and (2) no longer claim that we knew those things in the past, but we also (3) claim that our earlier claims to know them were false (Hawthorn, 2004; MacFarlane, 2005). We will take back earlier knowledge claims, even if we recognize that the context of attribution has changed. In this respect, ‘know’ seems quite different from words like ‘tall,’ which are more plausibly tied to their contexts of attribution. When I move from a third-grade to an adult context, I will no longer refer to my 10-year-old cousin as ‘tall,’ but I will not take back my earlier characterization of him as tall or admit that it was mistaken. I’d be more likely to flag the original context of attribution, saying something like, “I meant tall for a 10 year old,” than concede that my original claim was false. By contrast, when the skeptic challenges my knowledge claim, I’m not inclined to simply respond with something like, “I meant ‘know’ in the ‘easy’ sense.” Rather, if I find the skeptic’s arguments convincing, I’ll come to view my older knowledge claims as false.

Cases like this last one can be handled by adopting the more radical stance of claiming that the assessor’s context determines which standards are relevant for knowledge attributions (MacFarlane, 2005). Such assessor-based contextualism will lead to a much more robust relativism about knowledge. On such a view, relative to C’s context, not only will A not know

that his cell phone is in the room but B's claim that A does know this will also be false. In much the same way, relative to B's context, A knows that his cell phone is in the room, and C's claim that he doesn't know this is false. On this view, then, not only can a particular sentence ("John is tall" or "Peter knows that it is raining") be true when asserted in one context and false when asserted in another but also the same assertion of a sentence can be true relative to one context of assessment and false relative to another.

This last form of contextualism may do the best job of tracking our own use of epistemic terms, but the degree to which it requires us to complicate not only our general semantic framework but also our conception of truth may seem too radical for some. For these, it may seem more promising to simplify the semantics by trying to explain away those aspects of our practice that support the assessor-based view. Once again, however, more moderate contextualists will risk sliding to invariantism if these same strategies for explaining away part of our practice can be extended to the rest of those aspects of our use that conflict with invariantism.

See also: Analytic Philosophy; Epistemology and Language; Indexicality: Philosophical Aspects; Meaning: Overview of Philosophical Theories; Pragmatic Determinants of What Is Said; Truth Conditional Semantics and Meaning.

Bibliography

Annis D (1978). 'A contextualist theory of epistemic justification.' *American Philosophical Quarterly* 15, 213–219.

- Cohen S (1987). 'Knowledge, context and social standards.' *Synthese* 73, 3–26.
- Cohen S (1988). 'How to be a fallibilist.' *Philosophical Perspectives* 2, 581–605.
- Cohen S (1998). 'Contextualist solutions to epistemological problems: skepticism, Gettier, and the lottery.' *Australasian Journal of Philosophy* 76, 289–306.
- DeRose K (1992). 'Contextualism and knowledge attributions.' *Philosophy and Phenomenological Research* 52, 913–929.
- DeRose K (1995). 'Solving the skeptical problem.' *Philosophical Review* 104, 1–53.
- DeRose K (2002). 'Assertion, knowledge and context.' *Philosophical Review* 111, 167–203.
- Hawthorne J (2004). *Knowledge and lotteries*. Oxford: Oxford University Press.
- Lewis D (1996). 'Elusive knowledge.' *Australasian Journal of Philosophy* 74, 549–567.
- MacFarlane J (2005). 'The assessment-sensitivity of knowledge attributions.' *Oxford Studies in Epistemology* 1.
- Neta R (2002). 'S knows that p.' *Nous* 36, 663–681.
- Neta R (2003). 'Contextualism and the problem of the external world.' *Philosophy and Phenomenological Research* 66, 1–31.
- Rysiew P (2001). 'The context-sensitivity of knowledge attributions.' *Nous* 35, 477–514.
- Schaffer J (2004). 'Skepticism, contextualism, and discrimination.' *Philosophy and Phenomenological Research* 69, 138–155.
- Schiffer S (1996). 'Contextualist solutions to scepticism.' *Proceedings of the Aristotelian Society* 96, 317–333.
- Unger P (1984). *Philosophical relativity*. Minneapolis: University of Minnesota Press.
- Unger P (1986). 'The cone model of knowledge.' *Philosophical Topics* 14, 125–178.
- Williams M (1991). *Unnatural doubts*. Oxford: Blackwell.
- Wittgenstein L (1969). *On certainty*. Oxford: Blackwell.

Conventions in Language

M Kölbel, University of Birmingham, Birmingham, UK

© 2006 Elsevier Ltd. All rights reserved.

Independently of the question of what exactly linguistic meaning is (see **Meaning: Overview of Philosophical Theories**), a question arises as to the nature of its attachment to linguistic expressions: why does the word 'banana' mean what it does rather than something else? Why doesn't some other word have the meaning that 'banana' actually has? The answer almost everyone agrees upon is that it is a matter of convention that words mean what they do. Had there been different conventions of language, then words

would have had different meaning. Views diverge, however, on the significance of the conventionality of language, on the question of what exactly a convention of language is, and on the extent to which meaning is conventional (as opposed to, say, inferential). In what follows the focus will be mainly on the second of these issues, i.e., on the nature of linguistic conventions.

Convention and Analyticity

In the background of current thinking on language conventions is the attempt of the logical empiricists to

explain *a priori* knowledge as knowledge of analytic truths, i.e., propositions that are true in virtue of meaning (Carnap, 1947; Ayer, 1946; *see Analytic/Synthetic, Necessary/Contingent, and a Priori/a Posteriori: Distinction and A Priori Knowledge: Linguistic Aspects*). An example are the truths of arithmetic: while Kant had thought they were synthetic (not true in virtue of meaning), Ayer and Carnap followed Frege in claiming that they are analytic, i.e., true by definition. Carnap extended this approach to modality: necessary truths are just those that are true in virtue of linguistic rules.

Conventionalism was opposed by Quine, who argued against Carnap that there is no coherent way of drawing a distinction between analytic and synthetic truths (Quine, 1951, 1960). According to Quine, it is impossible to separate the conventional from the empirical ingredient of any truth, because every attempt to explicate analyticity will ultimately rely on some other inexplicable semantic notion, such as synonymy or possibility.

The debate between Carnap and Quine forms the historical background for recent efforts to explain in detail how language is conventional. The most influential account is that by David Lewis (1969, 1983), who provided a game-theoretic account of convention in general and then explained the specific nature of conventions of language within this framework. However, Lewis's account built on Grice's earlier analysis of linguistic meaning in terms of speaker intentions.

Grice

Grice claimed that linguistic meaning is ultimately a matter of the communicative intentions of speakers (Grice, 1989). Grice started by defining a notion of speaker meaning ('non-natural meaning') in terms of speaker intentions and then analyzed the meaning of expression types in terms of their use by speakers to speaker-mean something with them (*see Expression Meaning versus Utterance/Speaker Meaning*). He defined speaker-meaning as follows: a speaker *S* speaker-means that *p* by uttering *s* just if in uttering *s* *S* intends his or her audience to think that (*S* believes that) *p* on the basis of the audience's recognition of that very intention (Grice, 1989: 213–223, 123). For Grice, the meaning of expression types depended on what speakers in a speech community use these types to speaker-mean on particular occasions of use. A little more precisely, the dependence is as follows: a sentence type *s* means that *p* in a community *C* just if members of *C* have the habit of speaker-meaning that *p* by uttering *s*, and they retain the habit conditionally upon other members doing likewise. In short,

words mean what they do because speakers use these words habitually with certain communicative intentions, and this habitual procedure is conditional upon other speakers doing likewise. (For the fine details of the account, see Grice, 1989: 124–128.)

Lewis

Grice's analysis of linguistic meaning in terms of speaker intentions was initially perceived to be in competition with accounts offered by formal semanticists (see e.g., Strawson, 1969). The formal semanticist's central notion in the explanation of meaning is not that of intention but that of a 'truth condition.' However, it now seems that the two approaches can complement each other, and need not be viewed as competitors. Formal semanticists study artificial languages (which often serve as models of fragments of natural languages) with the aim of elucidating phenomena of compositionality (*see Compositionality: Philosophical Aspects and Formal Semantics*). Grice's framework does not address questions of compositionality, but it can in fact accommodate the formal semanticists' approach. David Lewis's theory of linguistic conventions not only showed how the insights of formal semantics can be appropriated within Grice's theory of communicative intentions; it also offered a detailed explication of the notion of convention itself (Lewis, 1969, 1983).

According to Lewis, there is a vast range of possible languages. Restricting himself initially to simple cases (languages with only context-insensitive declarative sentences), Lewis thought of a possible language as a function from a domain of sentences into a range of truth conditions. Many of the languages described by formal semanticists are possible languages in this sense. Most possible languages, however, are not used by anyone. According to Lewis, this is where convention plays a key role. He used his game-theoretic notion of convention to specify under what conditions a possible language is an actual language, i.e., is actually used by a population of language users.

Lewis's General Notion of Convention

Any word could in principle be used to mean anything. If two language users are to communicate successfully they therefore need to coordinate their use of words and make sure they use the same words with the same meaning. This type of situation, where several agents have a common interest in coordinating their actions, is called a 'coordination problem.' Conventions are a way of solving coordination problems – linguistic conventions are just a special case of this more general phenomenon.

According to Lewis, conventions (linguistic or not) are regularities in the behavior of the agents of a population. These regularities arise from the common interest of the agents to coordinate their actions and is sustained because each agent expects the others to conform to the regularity and prefers to conform him- or herself if the others conform. There are potential alternative regularities which could also secure coordination, hence the need for a convention.

For example, if our phone conversation is interrupted and we have the common aim of continuing the conversation, then there are two alternatives: either I phone back and you wait, or you phone back and I wait. No other combination of actions will achieve our common aim. Each of us prefers to phone back if the other waits and prefers to wait if the other phones back. But how do we know what the other is doing? If the problem is a recurrent one, then a convention can help. For example, if each of us expects the other to phone back just if the other was the original caller and not to phone back otherwise, then each of us will prefer to phone back if and only if he or she was the original caller.

Lewis's definition of convention is roughly as follows (see Lewis, 1983: 165 for full details): a regularity *R* is a convention in a population *P*, just if

1. everyone conforms to *R*;
2. everyone believes that the others conform to *R*;
3. the belief that the others conform to *R* gives everyone a decisive reason to conform to *R* him- or herself;
4. *R* is not the only regularity meeting (3);
5. (1)–(4) are common knowledge among *P*: they are known to everyone, it is known to everyone that they are known to everyone, etc.

Conventions of Language

Lewis used the above definition of convention in his explication of what it is for any of the many possible languages (as described by a semantic theory) to be the language actually used by a population. According to Lewis, a population *P* uses a possible language *L* just if members of *P* have a convention of uttering sentences of *L* only if they are true in *L*, and of coming to believe in the truth in *L* of sentences that are uttered by others. The relevant coordination problem for a population here is the problem of converging on one possible language. It is in the interest of each member to use the language the other members are using because there is a common interest in communication. Lewis called this a “convention of truthfulness and trust.”

There are some difficulties of detail that can be resolved by further refinements. For example, the proposal as sketched above does not take into account indexical languages or languages with non-declarative sentences (e.g., interrogative sentences). Lewis himself discussed how his approach can be suitably extended (Lewis, 1983). Another difficulty is the fact that too few speakers try to utter only sentences that are true in their language, and similarly too few speakers believe everything they are told. There is therefore no convention of truthfulness and trust in English among English speakers. Lewis's account can be modified to deal with this problem. For example, instead of saying that users of a language try to utter only sentences that are true in that language, Lewis could say that they utter sentences only if they **accept**, or want to commit themselves to, their truth for the purposes of the conversation.

A Basic Difficulty for Grice–Lewis

There are also some more fundamental difficulties, which concern the basic assumptions on which the Grice–Lewis approach is built. It is part of both Grice's and Lewis's accounts to attribute to language users highly complex mental states. On both accounts, language users are required to have unrealistically complex iterated preferences and beliefs concerning other language users (see definitions above). Typical language users, however, do not report these mental states. Lewis's response to these doubts concerning the psychological reality of these mental processes was to say that they are merely ‘potential’: users would explicitly have these cognitive states if they bothered to think hard enough (Lewis, 1983: 165) and presumably they would also be able to report these intentions if they thought hard enough. However, it is unclear whether the phrase ‘hard enough’ is substantial enough to render the theory empirically testable. Would Lewis accuse anyone denying the psychological reality of the account of not thinking hard enough?

Some psychological findings seem to add weight to this line of objection. The fundamental assumption behind Grice's and Lewis's approaches is that linguistic behavior is a product of a special case of instrumental reasoning. This much seems to be implied by Grice's idea that linguistic meaning is a matter of communicative intentions and linguistic behavior a special case of intentional action. As Laurence (1996) pointed out, however, there are cases which suggest that language processing and instrumental reasoning are independent faculties. A disability in instrumental reasoning

can be accompanied by full linguistic abilities. Conversely, lack of linguistic abilities can be accompanied by fully functioning instrumental reasoning.

Chomskyan Accounts of Linguistic Convention

A Chomskyan view of language processing lends itself to a different account of linguistic convention. Any account of linguistic convention needs to preserve the idea that what a given word means is a contingent and largely arbitrary matter; that words could have meant something other than what they actually mean, and that other words could have meant what they actually do. Laurence (1996) argued that a Chomskyan view does preserve this idea.

On such a Chomskyan view, language processing is performed by a special language-processing faculty. This faculty processes language at various levels, phonologically, syntactically, and semantically. At each level, the faculty associates certain representations with utterances. On this view, one might say that the various representations the language faculties of a group associate with a given utterance determine that utterance's meaning in the language of that group. The meaning of an expression type would then be a function of the representations the language faculties would associate with any utterances of that type.

On this view of the meaning of expression types, it does indeed turn out to be contingent: each type might have meant something other than it actually means, etc. For the precise working of the language faculty in an adult is partly the result of environmental influences. Within the constraints of universal grammar, children learn the language spoken in their surroundings. Thus, the representations computed by a given language faculty will depend in part on the language-learning environment. Had the environment been different, the representations associated by the language processor would have been different, thus its meaning would have been different.

This model works best for the conventions of a language spoken by people who have learnt the language in the natural way. But it would also explain explicit linguistic conventions (e.g., when a new technical term is explicitly introduced in a scientific paper, or when an adult learns a natural language). Presumably, these are cases where instrumental reasoning provides input for, and interacts with, the separate language-processing faculty.

Convention versus Inference

The controversy between Griceans and Chomskyans concerns the role of instrumental reasoning in the

determination of what expressions conventionally mean. There is another controversy, again involving Grice at centre stage, concerning the extent to which the meaning of utterances is the product of the conventional meaning of the expression types used as opposed to other, linguistically unanticipated, inferences. Grice distinguished between what is literally said by an utterance from what is 'implicated' (see **Semantics–Pragmatics Boundary**). What is literally said is more or less determined by the conventional meaning of the expressions used. However, language users often aim to convey messages that go beyond what is literally said, such as the polite referee in Grice's famous example: when the referee says "the candidate has an excellent command of English" he is relying on the audience's ability to infer that he wished to convey that the candidate is no good at philosophy (see Grice, 1989: 33).

The controversy concerns which aspects of communication should be viewed as arising from pragmatic inferences, as in the case of Gricean implicatures, and which aspects should be viewed as pertaining to literal meaning. (Another related question is whether any implicature can be conventional.) Davidson is at one end of the spectrum of possible views here: he practically denies (in good Quinean fashion) that there is any conventional meaning. It may be helpful in interpreting an utterance to start with a conjecture that the expression types uttered have certain stable meaning, but ultimately such a conjecture is merely a 'crutch' (Davidson, 1984: 279). (For more on these questions see Recanati, 2004; also **Semantics–Pragmatics Boundary** and **Nonstandard Language Use**).

See also: A Priori Knowledge: Linguistic Aspects; Analytic/Synthetic, Necessary/Contingent, and a Priori/a Posteriori Distinction; Compositionality: Philosophical Aspects; Expression Meaning versus Utterance/Speaker Meaning; Formal Semantics; Meaning: Overview of Philosophical Theories; Natural versus Nonnatural Meaning; Nonstandard Language Use; Semantics–Pragmatics Boundary.

Bibliography

- Ayer A J (1946). *Language, truth and logic* (2nd edn.). London: Victor Gollancz.
- Carnap R (1947). *Meaning and necessity*. Chicago: University of Chicago Press.
- Davidson D (1984). 'Communication and convention.' In Davidson D (ed.) *Inquiries into truth and interpretation*. Oxford: Oxford University Press. 265–280.
- Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Laurence S (1996). 'A Chomskian alternative to convention-based semantics.' *Mind* 105, 269–301.

Lewis D K (1969). *Convention*. Cambridge, MA: Harvard University Press.

Lewis D K ([1975] 1983). 'Languages and language.' In Lewis D (ed.) *Philosophical papers*, (2 vols). Oxford: Oxford University Press Vol. 1: 163–188.

Quine W V (1951). 'Two dogmas of empiricism.' *Philosophical Review* 60, 20–43.

Quine W V (1960). 'Carnap and logical truth.' *Synthese* 12, 350–374.

Recanati F (2004). *Literal meaning*. Cambridge: Cambridge University Press.

Strawson P F (1969). '*Meaning and truth*,' Inaugural Lecture, reprinted in *Logico-linguistic papers* (1971). London: Methuen. 170–189.

Cooperative Principle

K Lindblom, Stony Brook University, Stony Brook, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

The Principle Itself

In his William James Lectures at Harvard University in 1967, H. Paul Grice posited a general set of rules contributors to ordinary conversation were generally expected to follow. He named it the Cooperative Principle (CP), and formulated it as follows:

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged (Grice, 1989: 26).

At first glance, the Cooperative Principle may appear an idealistic representation of actual human communication. After all, as Grice himself has learned from his detractors, many believe "...even in the talk-exchanges of civilized people browbeating disputation and conversational sharp practices are far too common to be offenses against the fundamental dictates of conversational practice." Further, even if one discounts the tone of an exchange, "much of our talk exchange is too haphazard to be directed toward an end cooperative or otherwise" (Grice, 1989: 369).

However, Grice never intended his use of the word 'cooperation' to indicate an ideal view of communication. Rather, Grice was trying to describe how it happens that – despite the haphazard or even agonistic nature of much ordinary human communication – most discourse participants are quite capable of making themselves understood and capable of understanding most others in the course of their daily business.

What Counts as Cooperation?

Grice invites us to consider the following, quite unextraordinary exchange:

A: I am out of petrol.

B: There is a garage round the corner (Grice, 1989: 32).

Assuming A immediately proceeds to the garage, secures the petrol, and refills his car, we may describe B's contribution as having been successful. By what rational process of thought was A so quickly able to come to the conclusion that the garage to which B refers would fulfill his need for petrol? Why did B's utterance work? Grice's answer: because A and B adhere to the Cooperative Principle of Discourse.

It is not hard to imagine that two friends sharing a ride would want to help each other through a minor crisis; thus, 'cooperation' in this scenario seems quite apt.

But imagine the exchange went this way instead:

A: I am out of petrol.

B: (sarcastically) How nice that you pay such close attention to important details.

In this second scenario, not only does B refuse to assist A in solving the problem, he uses the occasion to add to A's conundrum an assault upon his character. Assuming A feels the sting, again B's contribution has been successful. So how and why in this case has B's contribution worked? How can such a sour response as B's callous retort be considered 'cooperative'? Again, Grice's Cooperative Principle proves a useful answer. The explanation requires closer inspection of the strictness with which Grice uses the term.

The Cooperative Principle and the Maxims of Cooperative Discourse

Grice explicates his Cooperative Principle of Discourse in 'Logic and Conversation,' the paper originally presented at Harvard University in 1967, later printed in Cole and Morgan (1975), and reprinted in a slightly revised version in Grice's *Studies in the Way of Words* (1989). We cite from his final version as we assume this is the one he considered most complete. In the essay, Grice is careful to limit use of the CP for describing only talk exchanges that exhibit the following three specific characteristics:

1. The participants have some common immediate aim.
2. The contributions of the participants [are] dovetailed, mutually dependent.

3. There is some sort of understanding (often tacit) that, other things being equal, the transactions should continue in appropriate style unless both parties are agreeable that it should terminate (Grice, 1989: 29).

Though he is careful to limit the CP's application to talk exchanges that exhibit these particular cooperative characteristics, this list should not be read as an admission of great limitation. For Grice finds that most talk exchanges **do** follow the CP because most talk exchanges do, in fact, exhibit the **cooperative** characteristics he outlines:

Our talk exchanges ... are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction (Grice, 1989: 26).

Grice identified the Cooperative Principle as a 'super principle' or a 'supreme principle' (1989: 368–369) that he generalized from four conversational 'maxims' he claimed discourse participants ordinarily follow. With a nod to Kant, Grice identifies the maxims as:

1. Quantity (give as much information as is required, and no more than is required)
2. Quality (do not say what is false or that for which you lack adequate evidence)
3. Relation (be relevant)
4. Manner (be clear, be orderly, and avoid ambiguity) (1989: 28).

Clear fulfillment of these maxims may be demonstrated in the following exchange:

A: Do you know where I can buy some petrol?
B: You can buy petrol at the garage right around the corner.

Let us assume that B is sincere and knowledgeable, and A finds the garage right away based upon B's advice. It is the case then that B's response to A's question follows the maxims completely, giving exactly the right amount of information (quantity), information for which B has the required evidence (quality), information that is directly connected to A's question (relevance), and information given in a fashion effectively and efficiently understood (manner).

But Grice knew that people do not always follow these maxims as they communicate. (What dull business conversation analysis would be if they did!) Rather, interlocutors can fail to fulfill the maxims in a variety of ways, some mundane, some inadvertent, but others lead to what most consider the most powerful aspect of Grice's CP: conversational 'implicature.'

Failures to Fulfill Maxims and Implicature

Grice describes four ways in which maxims may go unfulfilled in ordinary conversation. The first three ways are fairly straight forward. One might **violate** or **infringe** a maxim. This infringement is often done with the intention of misleading; for example, one might say, 'Patricia was with a man last night' as a way of making Patricia's routine dinner out with her husband seem clandestine. One might **opt out**, making it clear that one refuses to cooperate in a conversation for some reason; for example, one may be legally bound not to provide information one has. Or, one might encounter a **clash** of maxims, facing the choice of violating one maxim or another. For example, one may not be able to give all of the information required (quantity) because one does not have adequate evidence for the information (quality).

Most interesting is the final possibility for the non-fulfillment of a maxim: **flouting** or **exploiting** a maxim for the purpose of implicating information (implicature). This case is the one in which even an apparently uncooperative response illustrates discursive or linguistic cooperation. Recall the examples with which this article was introduced.

A: I am out of petrol.
B: There is a garage round the corner.

In this instance, we may claim, that B – at first blush – appears to break the maxim of relation. For what does a garage have to do with petrol? Since drivers are aware that garages sell petrol, it is not long before A realizes that B has not broken the maxim of relation at all; it is, in fact, instantaneous. B's point is directly relevant. B is being cooperative in both the colloquial sense and the specialized sense Grice applies to the term. Grice's Cooperative Principle makes sense of the speed with which A is able to process the usefulness of B's contribution. A assumes B is following the maxims and would thus not mention the garage unless it had petrol.

In the next scenario, however, the exchange, and thus the rational process by which A makes sense of B's contribution, is markedly different:

A: I am out of petrol.
B: (sarcastically) How nice that you pay such close attention to important details.

In this instance, B flouts the maxim of quality by stating as true something for which he has specific and immediate evidence is untrue. One likely implication of B's remark is that A is an idiot for not paying attention to such an important detail as having enough petrol in the car. If A feels the sting of B's remark, A and B have exhibited discursive

cooperation that resulted in an implicature directed to A from B (see **Maxims and Flouting**).

While one example hardly illustrates so many cases, Grice works out a number of possible forms of implicature: irony, metaphor, meiosis (understatement), hyperbole, social censure, deliberate ambiguity, and deliberate obscurity (for example, if one is trying to keep a secret from the children). In all of these cases, maxims are broken and the breaks result in specific information implied to and understood by the receiver of the utterance.

The power of the conversational maxims to describe rational processes by which speakers and hearers make sense of each other's utterances have energized many scholars of language and conversation across many fields. But, as the introduction to this article makes clear, the Cooperative Principle has not been free from serious critique.

Major Critiques of the Cooperative Principle

Problems with the Term 'Cooperation'

Despite the care with which he used the term "cooperation," Grice is regularly accused of promulgating a theory that assumes too friendly a spirit of communicative interaction among people. This charge is most commonly made in work outside of Grice's own field of linguistic philosophy. In effect, these detractors claim Grice is just too nice.

For example, Tannen (1986) claims that Grice's maxims of cooperative discourse can't apply to "real conversations" because in conversation "we wouldn't want to simply blurt out what we mean, because we're judging the needs for involvement and independence" (1986: 34–45). Tannen assumes that Grice's maxims are prescriptions that conversations must follow strictly in order to be considered cooperative. Cameron (1985) makes a similar case, taking issue with Grice's application of the term 'cooperation' to all discourse. Cameron is quite correct in her claim that – at least in the colloquial sense of the term – assumptions regarding the appropriateness of 'cooperative' behavior have dogged women for centuries. But Cameron demonstrates a reductive view of Grice's use of the term 'cooperation' when she describes Grice's CP as an 'inflexible' and 'unproductive' apparatus that provides yet another way for both 'chauvinists and feminists' to believe that 'whereas men compete in competition, women use co-operative strategies' (1985: 40–41). Grice's version of cooperation is more flexible and less dogmatic than these critics assume.

Others have gone so far as to claim Grice **advocated** cooperation among conversational participants,

believing Grice prescribed cooperation as the most effective way of engaging in meaningful communication with others.

Cooper (1982), interested in applying Grice to theories of written composition, claims that Grice advocates cooperation because

what enables conversation to proceed is an underlying assumption that we as conversants have purposes for conversing and that we recognize that these purposes are more likely to be fulfilled if we cooperate (1982: 112).

The notion that discourse participants cooperate with each other and that they do so out of a mutual benevolence is a misreading of Grice's position on cooperative discourse; but it is one that persists.

Grice himself acknowledged the difficulty some have had interpreting his use of 'cooperation.' As a final chapter to his 1989 book, Grice wrote a 'Retrospective Epilogue' in which he considered criticism his theories had engendered. It has already been related that here Grice acknowledged that his theory suffers from a perceived naïveté. To combat the criticism, Grice adds useful information about what counts as cooperative in discourse. First, he reminds readers of the sort of utterances he seeks to elucidate: voluntary talk exchanges that require some form of "collaboration in achieving exchange of information or the institution of decisions." And, he points out that within exchanges intended to produce information or determine decisions, cooperation "may coexist with a high degree of reserve, hostility, and chicanery and with a high degree of diversity in the motivations underlying quite meager common objectives" (Grice, 1989: 369). Even as adversarial an exchange as a hostile courtroom cross-examination would at least simulate adherence to the CP.

To further explain the sort of cooperation to which Grice refers, it might help to borrow a term from classical rhetoric. The ancient Greeks used the term 'Nomos' to indicate cultural practices that defined a group of people. Two closely related connotations of the term are useful for the present discussion: (1) 'the mores' of a given collective (Ostwald, 1969: 33); and, (2) customs "which are generally observed by those among whom they prevail" (1969: 36). Nomos is not necessarily an explicit, prescribed set of conventions, but rather a set of conventions that are brought into existence by the very fact that people ordinarily follow them, perhaps without even realizing they are following a set of conventions. When American youths visit Europe, the locals can spot them in an instant by their footwear; but, in the United States, sneakers are simply what young people ordinarily wear.

Nomos applied to conversation, then, is a set of conventions, or rules (or maxims) for talk according to

which a group of people ordinarily makes meaning. In the maxims, Grice believes he has found universal conventions that all people may regularly follow in their meaning-making talk exchanges. In order for such a set of conventions to function, a certain degree of at least tacit assent to those conventions is necessary. Thus, the term ‘cooperation’ is quite apt.

The crucial subtlety of Grice’s theory is this: interlocutors do not necessarily cooperate with each other; they cooperate with a set of conventions that allows each interlocutor to produce approximate enough meanings for communication to work. This form of cooperation is not necessarily benevolent at all; even the bitterest of verbal fights require linguistic cooperation to work.

The aim for Gricean conversation analysis – and thus the CP and the maxims – is not to advocate benevolent cooperation, but to prove the rationality of conversation. “...observance [of the maxims] promotes and their violation [except in the case of implicature] dispromotes conversational rationality” (Grice, 1989: 370).

Although many have claimed Grice’s writing on the CP is ambiguous and is on occasion inconsistent with terminology, this should not be said of Grice’s measured use of the term ‘cooperation.’

Precise readings of Grice’s writing on cooperation demonstrate that he rarely, if ever, describes interlocutors as **being** cooperative. Rather, he claims that interlocutors’ contributions to conversation are **cooperative**. The contributions are uttered in cooperation with a set of conventions for producing meaning. In this sense, we might think of a pair of interlocutors as each operating according to the dictates of a set of conventions (the maxims) and thus they are ‘co/operators’: two operators of discourse operating at once.

Consider also, Grice’s use of the term ‘dovetailed’ in describing the state of cooperative contributions to conversation (1989: 29). Dovetailed elements are placed within very close proximity to each other, maintaining the integrity of each separate element, but creating a stronger whole. Utterances remain utterances, but conversations take flight, implicating new meaning for hearers and speakers.

Problems with the Maxims: The Haphazardness of Communication and the Specificity of Maxims

The second major critique of the Cooperative Principle has been a topic of spirited discussion among linguistic philosophers since Grice first proposed it. Grice himself identifies the problem as resulting from the thought that communication is simply too “haphazard” to be described accurately as having a cooperative end. Some forms of communication are not

appropriately described by the CP. For example, as Grice puts it, “Chitchat goes nowhere, unless making the time pass is a journey” (1989: 369).

Grice suggests the problem is two-fold. First, he agrees with critics that the maxims appear less “coordinate” than he would prefer. The maxim of quality appears in some ways more definitive of information than the other maxims. And, the maxims are not independent enough: relevance, as will be shown, has been often regarded as containing the essence of the other maxims. Second, Grice’s selection of cooperation as the “supreme Conversational Principle” underpinning the rationalizing operations of implicature remains, to say the least, not generally accepted (1989: 371).

In his ‘Conversational maxims and rationality’ Kasher (1976), claims that cooperation is not a principle that accounts for all information conveyed by implicature because cooperation may be “contrary to [a speaker’s] interest” (1976: 241). Kasher offers the following example: Man A. is asked by Man B. “Who is going to marry your sister?” Man A., who knows the proper name of the intended, replies, “A peacock dealer.” Man A.’s reply, Kasher points out, does not satisfy the demands of full cooperation, and the CP, claims Kasher, cannot account for a situation in which there is no cooperation. As an alternative explanation for the operation of conversational implicature, Kasher poses the “Rationalization Principle,” which stems from the idea that Relevance (one of Grice’s maxims) is the only necessary element to explain a talk exchange. In a later work, Kasher renames his principle “the principle of rational coordination,” which states: “Given a desired basic purpose, the ideal speaker chooses that linguistic action which, he believes, most effectively and at least cost attains that purpose” (Kasher, 1977). Kasher’s well known critique thus began what has become ‘Relevance Theory,’ which is at its base a refinement of Grice’s earlier work (*see Relevance Theory*). (See below for references to other work on Relevance.)

Though in his final work he admitted some misgivings and offered minor refinements of his maxims of cooperative discourse, Grice, up until his death in 1988, defended his selection of the Cooperative Principle as the ‘supreme principle.’

Scholarship Influenced by the Cooperative Principle

Though critiques of the CP remain unresolved – and perhaps they always will be – there is nevertheless no denying that Grice’s CP has had a dramatic influence on discourse studies across disciplines. The CP can

probably not be considered definitive, but there is no denying it has proven quite generative.

Because Grice's Cooperative Principle has such cross-disciplinary appeal, any survey of work influenced by it is almost certainly incomplete. The sketch here is intended to acquaint the reader with some applications of major importance and to give readers a richer understanding of the depth and breadth of the influence Grice has had. (For more citations and commentary on work influenced by Grice's CP, see Lindblom, 2001.)

Grammar

Grammarians frequently view literal or sentence meaning as more important than any individual's intended meaning in making an utterance. Thus Chomsky, for example, has critiqued Grice's CP for being unprincipled (1975: 112) and has complained that Grice's approach to language study is behaviorist due to his focus on utterer's intention (Suppes, 1986: 121). Other grammarians influenced by Chomsky have used similar logic to critique the CP as too concerned with context.

Suppes, whose essay is an excellent synthesis of grammar studies using Grice, argues that these grammarians assume an even more closely rule-bound language governance, making their claims essentialist. Further, he argues, that Grice's CP is useful precisely because it is so context dependent. Chomsky's positivism is not an issue in a Gricean analysis because Grice's work "bring[s] out the importance of context" (Suppes, 1986: 124).

Neo-Gricean Pragmatics

Grice's influence is most apparent in a branch of linguistic study that has become known among some as Neo-Gricean pragmatics. Scholars in this field have greatly revised Grice's maxims of cooperative discourse in a variety of interesting ways, but they have maintained the basic direction of Grice's work, especially in regard to the concept of conversational implicature. Huang (1991) usefully surveys a great deal of scholarship from well known scholars in this area, including Atlas, Levinson, Sperber and Wilson, Leech, and Horn.

As mentioned previously, Kasher developed a specific focus on one of Grice's maxims, thus establishing the field of Relevance Theory. Sperber and Wilson have also generated an important Relevance Theory, theirs influenced by Fodor's theory of cognitive modularity. According to Huang, Sperber and Wilson believe "one is always maximizing the informational value of contextual stimuli to interpret the utterance in a way which is most consistent with

the Principle of Relevance" (Huang, 1991: 303). Along with texts by Kasher and Sperber and Wilson, important developments in Relevance Theory may also be found in Grandy and Warner (1986) and Tsohatzidis (1994).

More recently, a special issue of *Journal of Pragmatics* has focused exclusively on Gricean themes in pragmatic analysis. Although he resists the notion of a school of 'Neo-Gricean' approaches, the journal editor has nevertheless gathered a collection of papers that illustrates that Grice's CP and maxims are ideas that "shook the world of language study in the past century, and continue to move and inspire today's research" (Mey, 2002: 911). The special issue includes essays focused on social roles in Japan, maxim confluence among multi-lingual code-switchers, academic writing, and other current approaches to Gricean pragmatics.

The CP is not only applicable across cultures, it is also possible to use Gricean analysis to examine a 'theme' in discourse. For example, much interesting work is underway in the pragmatics of humor (for example, Attardo, 2003).

Politeness Theory

Politeness theorists use Grice's CP specifically to examine the ways in which maxims are exploited to indicate some special status of the hearer. For example, a lawyer would answer a judge, "Yes, your honor." The 'your honor' breaks the maxims of quantity – as surely the judge is aware of her title – but including the words 'your honor' implies the speaker's understanding that the judge holds a greater position of authority.

For a valuable survey of Politeness Theories, see Fraser (1990). In this piece Fraser examines politeness theories posited by Lakoff and by Leech, and he explains that both of these theories rely heavily on Grice's CP, though Lakoff reduces the maxims by two and Leech increases the number by six. The most influential Politeness Theory was developed by Brown and Levinson (1987).

Brown and Levinson's work is primarily influenced by Goffman, but they also claim "Grice's theory of conversational implicature and the framework of the maxims that give rise to such implicatures is essentially correct" (1987: 3). Goffman's influence may be seen in Brown and Levinson's concentration on the concept of 'face wants.' Their politeness theory examines the ways in which speakers and hearers use conversational implicature to fulfill the 'face wants' of higher-status participants in conversation. Like the CP itself, Politeness Theory is certainly not free from critique, but it has resulted in fascinating analysis and has generated much spirited debate.

Question Processing

Several works in the area of question processing have developed from Grice's Cooperative Principle. Questions and questioning patterns can result in implicatures regarding politeness, status, and authority, and they operate according to conventions that many have build upon Grice's maxims. Singer provides a useful assessment of the study of question processing in all of its stages: question encoding, question categories, selection of answering strategies, memory search, comparison, and response (1990: 261). He identifies 'response' as the category for which Grice's CP is the most powerful.

Most interesting in 'response' is Lehnert's theory of secondary questions. According to Singer, "If asked 'Are there oil wells in Manitoba?' a simple 'no' would appear rather blunt. Instead, in keeping with Grice's 'maxim of quantity' and Lehnert's theory of secondary questions, it is more appropriate to hypothesize the next logical question and answer, 'There are a few, but there is not much oil east of Saskatchewan.'" (Singer, 1990: 273).

Gender Studies

Though above we single out some scholarship in gender studies for applying superficial accounts of the CP, there is excellent scholarship in the field that has used Grice's CP and maxims to examine behavioral and status differences between women and men. Brown, using the Politeness Theory she developed with Levinson, has used Grice to examine the socio-political situations of women in non-Western cultures (1990). Rundquist and Michell have looked at men's and women's use of conversational strategies in western culture.

Rundquist uses Grice to confront the "popular belief that women's speech is more indirect than men's" (1992: 431). She finds that men more frequently than women flout maxims to implicate information. Some of the purposes she identifies for which men tend to implicate information include to "give direction to their children," to "put themselves down as well as to tease others," "to be humorous," "to show themselves off to their best advantage in conversation," and perhaps most significantly for a study of gender, "to avoid direct confrontation" (Rundquist, 1992: 447). Michell (1984) questions if women often flout maxims to implicate information. She determines that women are far more likely to simply lie to protect themselves from verbal and physical abuse in a misogynist culture. For example, imagine a woman missed a meeting because she had painful menstrual cramps and because she had an important report to finish. This woman would be far more

likely to claim she missed the meeting because of the report, leaving out the mention of cramps, even if the report was not even close to being the primary reason for her absence; her omission is an opting out, not an implicature (Michell, 1984: 376).

Teacher Research and Pedagogy

Studies in teacher research have approached Grice's Cooperative Principle for two important purposes: (1) to examine the discourse of the classroom situation; and (2) to establish effective pedagogical strategies.

Three valuable works serving the first purpose may be found in Edwards and Mercer (1987), Kleifgen (1990), and McCarthy (1987). The first two works focus closely on the ways in which the educational scenario highlights the need for listeners to fill in propositions implicated by speakers. Edwards and Mercer examine the ways in which children become more and more proficient in these skills through their educational training. Kleifgen suggests that teachers should look for the points in classroom discourse when students begin to predict the outcomes of teachers' questions so quickly that it is clear the students are ready to move on to a higher level of difficulty.

McCarthy's essay – probably the finest treatment of the CP from a pedagogy scholar – traces the development of a college student as he writes for his composition, cell biology, and poetry classes. Examining both the student's written assignments and the teachers' detailed responses to them, McCarthy uses Grice's CP to determine what is required for this student to cooperate as a writer in each class and whether or not he was successful. In McCarthy's judgment, the student was successful as a student because he was able to determine "what counted as 'cooperation'" in each of his classes (1987: 249). Thus, McCarthy uses the CP in a flexible, context specific manner consistent with Grice's own descriptions of it.

Other scholars with an interest in writing instruction have used Grice for productive ends. Though they are too likely to read Grice's CP as describing a benevolent, cooperative relationship between writer and reader, Cooper (1982; 1984) and Lovejoy (1987) have used the CP to positive effect in college writing classes. Lovejoy's very practical revising template using the maxims is especially useful for college students learning to write more sophisticated texts.

Professors of literature have also found Grice's CP of use in articulating abstract themes from literature. Pratt's (1977) work is probably the best known, but for a fascinating reading of Beckett's *Waiting for Godot* using Gricean analysis, see Gautam and Sharma (1986).

Conclusion

A cross-disciplinary examination of how Grice's Cooperative Principle has been put into practice clearly indicates that the CP has had tremendous appeal and influence. It is precisely the CP's flexibility and context-dependent nature that makes it of such broad value. However, that same flexibility and context-dependence has also generated a fair number of critiques that cite lack of specificity and a too-relativistic application to discourse. Thus, it seems, the CP's strength is also its weakness. Certainly a great diversity of scholars have found the Cooperative Principle of Discourse and its attendant Maxims of Conversational Cooperation useful as analytical tools toward a variety of ends. It is doubtful, however, that the notion of 'cooperation' among discourse participants will ever be universally accepted.

See also: Maxims and Flouting; Relevance Theory.

Bibliography

- Attardo S (2003). 'Introduction: the pragmatics of humor.' *Journal of Pragmatics* 35, 1287–1294.
- Brown P (1990). 'Gender, politeness, and confrontation in Tenejapa.' *Discourse Processes* 13, 123–141.
- Brown P & Levinson S C (1987). *Politeness: some universals in language use*. Cambridge: Cambridge University Press.
- Cameron D (1985). *Feminism and linguistic theory*. New York: St. Martins.
- Chomsky N (1975). *Reflections on language*. New York: Pantheon.
- Cooper M M (1982). 'Context as vehicle: implicatures in writing.' In Nystrand M (ed.) *What writers know: The language, process, and structure of written discourse*. New York: Academic Press. 105–128.
- Cooper M M (1984). 'The pragmatics of form: how do writers discover what to do when?' In Beach R & Bridwell L S (eds.) *New directions in composition research*. New York: The Guildford Press. 109–126.
- Edwards D & Mercer N (1987). *Common knowledge: the development of understanding in the classroom*. London: Routledge.
- Fraser B (1990). 'Perspectives on politeness.' *Journal of Pragmatics* 14, 219–236.
- Gautam K & Sharma M (1986). 'Dialogue in *Waiting for Godot* and Grice's concept of implicature.' *Modern Drama* 29, 580–586.
- Grandy R & Warner R (eds.) (1986). *Philosophical grounds of rationality: intentions, categories, ends*. Oxford: Clarendon.
- Grice H P (1975). 'Logic and conversation.' In Cole P & Morgan J L (eds.) *Syntax and semantics: Speech acts 3*. New York: Academic Press. 58–85.
- Grice H P (1989). *Studies in the way of words*. Cambridge: Harvard University Press.
- Huang Y (1991). 'A neo-Gricean pragmatic theory of anaphora.' *Journal of Linguistics* 27, 301–335.
- Kasher A (1976). 'Conversational maxims and rationality.' In Kasher A (ed.) *Language in focus: foundations, methods and system*. Reidel: Dordrecht. 197–216.
- Kasher A (1977). 'Foundations of philosophical pragmatics.' In Butts R E & Hintikka J (eds.) *Basic problems in methodology and linguistics: part three of the proceedings of the Fifth International Congress of Logic, Methodology and Philosophy of Science, London, Ontario, Canada-1975*. Reidel: Dordrecht. 225–242.
- Kleifgen J A (1990). 'Prekindergarten children's second discourse learning.' *Discourse Processes* 13, 225–242.
- Lehnert W (1978). *The process of question answering*. Hillsdale, New Jersey: Erlbaum.
- Lindblom K (2001). 'Cooperating with Grice: a cross-disciplinary metaperspective on uses of Grice's Cooperative Principle.' *Journal of Pragmatics* 33, 1601–1623.
- Lovejoy K B (1987). 'The Gricean model: A revising rubric.' *Journal of Teaching Writing* 6, 9–18.
- McCarthy L P (1987). 'A stranger in strange lands: a college student writing across the curriculum.' *Research in the Teaching of English* 21, 233–265.
- Mey J (2002). 'To Grice or not to Grice.' *Journal of Pragmatics* 34, 911.
- Michell G (1984). 'Women and lying: a pragmatic and semantic analysis of "telling it slant."' *Women's Studies International Forum* 7, 375–383.
- Ostwald M (1969). *Nomos and the beginnings of the Athenian democracy*. Oxford: Oxford University Press.
- Pratt M L (1977). *Toward a speech act theory of literary discourse*. Bloomington: Indiana University Press.
- Rundquist S (1992). 'Indirectness: a gender study of flouting Grice's maxims.' *Journal of Pragmatics* 18, 431–449.
- Singer M (1990). 'Answering questions about discourse.' *Discourse Processes* 13, 261–277.
- Sperber D & Wilson D (1986). *Relevance: communication and cognition* (2nd edn. 1995). Oxford: Basil Blackwell.
- Suppes P (1986). 'The primacy of utterer's meaning.' In Grandy R & Warner R (eds.) 109–130.
- Tannen D (1986). *That's not what I meant: how conversational style makes or breaks your relations with others*. New York: Morrow.
- Tsohatzidis S (ed.) (1994). *Foundations of speech act theory: philosophical and linguistic perspectives*. London: Routledge.

Coreference: Identity and Similarity

Y Huang, University of Reading, Reading, UK

© 2006 Elsevier Ltd. All rights reserved.

Defining Coreference

Coreference can in general be defined as the phenomenon whereby two or more anaphoric or referential expressions denote, or refer to, the same entity in the external world. This can be illustrated by (1).

- (1) John said that he had won two Olympic gold medals.

In (1), if the anaphoric pronoun *he* refers to what its antecedent *John* refers to, then there is a relation of coreference obtaining between them; *he* is thus said to be coreferential with *John*. In contemporary linguistics, including generative grammar, coreference between two or more anaphoric or referential expressions in a sentence or discourse is usually marked by using identical subscript letters, as in (2) or numbers, as in (3).

- (2) John_i said that he_i had won two Olympic gold medals.
(3) John₁ said that he₁ had won two Olympic gold medals.

Another common way of saying that *he* and *John* are coreferential is to say that they are coindexed. On the other hand, in (1) if *he* does not refer to what *John* refers to, then the two anaphoric/referential expressions are disjoint in reference. Disjoint in reference is typically marked by using different subscript letters, as in (4), or numbers, as in (5).

- (4) John_i said that he_j had won two Olympic gold medals.
(5) John₁ said that he₂ had won two Olympic gold medals.

Identity

From a truth-conditional, semantic point of view, the anaphoric relation exhibited in (1) is called referential anaphora (e.g., Huang, 2000: 5). A referential anaphoric expression refers to some entity in the external world either directly, as in (6), or via its coreference with its antecedent in the same sentence or discourse, as in (1). In the latter case, as already mentioned, the referentially anaphoric expression refers to what its antecedent refers to, thus they are of referential identity.

- (6) (Referent in the physical context, and with selecting gesture)
He's the robber!

Similarity

We move next to types of anaphoric relations that may be said to indicate some kind of anaphoric or referential dependency other than referential identity.

Bound-Variable Anaphora

Sentence (7) is an example of bound-variable anaphora.

- (7) Every child₁ wishes that he₁ could visit the land of Lilliput.

Generally speaking, a bound-variable anaphoric expression does not refer to any fixed entity in the external world, as can be shown by sentences such as (8) below.

- (8) Nobody₁ thought that he₁ would wish to live with the giants of Brobdingnag.

But it is interpreted by virtue of its dependency on some quantificational expression in the same sentence or discourse, thus seeming to be the natural language counterpart of a bound variable in first-order logic. As indicated by (7) and (8), the bound-variable anaphoric expression and its quantificational antecedent can be coindexed, but they are not considered to be coreferential.

One interesting characteristic of bound-variable anaphora is that different languages afford their speakers different types of anaphoric or referential expressions to encode such a dependency, (cf. Huang, 2000: 6). For example, to express a bound-variable anaphoric relation between a matrix subject and an embedded subject, while English normally allows neither gaps (or empty categories) nor reflexives, Serbo-Croatian allows gaps (or empty categories), Marathi allows reflexives, and Chinese allows both.

- (9a) Gaps or empty categories (Serbo-Croatian cited in Huang, 2000: 6)
svaki student misli da ce Ø
every-M-SG student thinks that will
dobiti desetku.
get A
'Every student thinks that he will get an A.'
- (9b) Pronouns
Every actress said that her career has been a roller coaster ride.
- (9c) Reflexives (Marathi, cited in Huang, 2000: 6)
sarvāānaa₁ vaatta ki aapan₁ libral aahot.
everybody believes that self liberal is
'Everybody believes that he is liberal.'

- (9d) Gaps and reflexives (Chinese)
- | | | | | | | |
|--------|----|--------------|-----|------|----|------|
| mei | ge | ren | dou | shou | Ø/ | ziji |
| every | CL | person | all | say | | self |
| xihuan | | zhongguocai. | | | | |
| like | | Chinese food | | | | |
- ‘Everybody says that he likes Chinese cuisine.’

Note next that crosslinguistically bound-variable anaphora occasionally can also be encoded by repeating the same lexical NP.

- (10) Of every ritual bronze that was found in the tomb, it was subsequently discovered that the bronze belonged to the Chinese Northern Song élite.

Finally, as noted in Kempson (1988) and Huang (1994: 292), examples of the following kind can also have a bound-variable interpretation. On such a reading, the supervisor is interpreted as each Ph.D. student’s supervisor. Of particular interest here is that this bound-variable interpretation is obtained only by virtue of the addition of the pragmatic inference that every Ph.D. student characteristically has a supervisor.

- (11) Every Ph.D. student thinks that the supervisor is intelligent.

E-Type Anaphora

Somewhat related to bound-variable anaphora is E-type anaphora, also known as *donkey* anaphora, first discussed by Geach (1962: 128). It is called E-type anaphora in memory of the late Oxford philosopher Gareth Evans after Evans (1977). A classical example of E-type anaphora is given in (12).

- (12) Every farmer who owns a donkey beats it.

For technical reasons, an E-type anaphoric relation is neither pure referential anaphora nor pure bound-variable anaphora, but appears to constitute a unified semantic type of its own (Evans, 1977). The main reason why it is neither pure referential anaphora nor pure bound-variable anaphora is this: the antecedent of the anaphoric pronoun *a donkey* is variable-bound by a quantificational expression *every farmer*, but unlike in the case of pure bound-variable anaphora such as (7), the antecedent does not syntactically bind the anaphoric pronoun, because the antecedent does not c-command the anaphoric pronoun. Put differently, in E-type anaphora, the anaphoric expression falls outside the scope of its binder (see also Heim, 1990; Kamp and Reyle, 1993; de Swart, 1998: 127–130).

Anaphora of Laziness

The following is a classical example of anaphora or pronoun of ‘laziness’ (e.g., Karttunen, 1976).

- (13) The man who gave his paycheck to his wife was wiser than the man who gave it to his mistress.

Anaphora of laziness is so-called because it exhibits neither a referential anaphoric relation nor a bound-variable anaphoric relation. Rather, it functions as a shorthand for a repetition of its antecedent, which supplies the descriptive content for the anaphoric expression. In other words, it is a device for a repeated occurrence of the linguistic form, rather than the truth-conditional content of its antecedent. This can be illustrated by a consideration of (13). In (13), the referent of *it* is the paycheck of the second man rather than the paycheck of the first man. There is thus no coreferential relation between *it* and *the paycheck*. Anaphora of laziness is considered as a case of the semantically defined type of identity of sense anaphora, that is, anaphora in which the anaphoric expression and its antecedent are related in terms of sense. It is on a par with N-bar-anaphora, as in (14), and arguably with the sloppy reading of VP-ellipsis, as in (16b) (see Huang, 2000: 131–156 for further discussion of VP-ellipsis).

- (14) Mary’s favorite tenor is Pavarotti, but Jane’s Ø is Carreras.

- (15) John loves his wife, and Peter does, too

- (16a) Strict reading
John likes John’s wife, and Peter loves John’s wife.

- (16b) Sloppy reading
John likes John’s wife, and Peter loves Peter’s wife.

Identity of sense anaphora contrasts with the semantically defined type of identity of reference anaphora, that is, anaphora in which the anaphoric expression and its antecedent have identical referent, as (2) illustrates.

Bridging Cross-reference Anaphora

Bridging cross-reference anaphora is used to establish an association with some preceding expression or antecedent in the same sentence or discourse via the addition of background assumption (e.g., Clark, 1977; Huang, 1994, 2000). What is tacitly bridged is typically the information that is not structurally retrievable either from the sentence or discourse that triggers the inferential process. A typical example of bridging cross-reference is given below.

- (17) John walked into a church. The stained glass windows were magnificent.

In (17), the anaphoric expression is *the stained glass windows*, and its antecedent is *a church*. The pragmatically inferred background assumption is that the church John walked into has stained glass windows.

As pointed out in Huang (2000: 249), bridging cross-reference anaphora has three characteristic properties:

1. The anaphoric expression, which is usually a definite NP, must occur in the appropriate context of its antecedent, which is usually an indefinite NP,
2. There is some semantic and/or pragmatic relation between the anaphoric expression and its 'antecedent,'
3. The anaphoric expression and its antecedent do not stand in a strictly coreferential relation. Rather, they are linked to each other via the addition of some pragmatic inference (see Huang, 2000: 249–253 for further discussion).

Other interesting cases of anaphoric or referential dependency other than referential identity may include split antecedence and overlap in reference.

See also: Anaphora: Philosophical Aspects; Deixis and Anaphora: Pragmatic Approaches; Donkey Sentences; Reference: Philosophical Theories.

Bibliography

- Clark H (1977). 'Bridging.' In Wason P & Johnson-Laird P (eds.) *Thinking: readings in cognitive science*. Cambridge: Cambridge University Press. 411–420.
- De Swart H (1998). *Introduction to natural language semantics*. Stanford. CSLI.
- Evans G (1977). 'Pronouns, quantifiers, and relative clauses. Parts I and II.' *Canadian Journal of Philosophy* 7, 467–536.
- Geach P (1962). *Reference and generality*. Ithaca: Cornell University Press.
- Heim I (1990). 'E-type pronouns and donkey anaphora.' *Linguistics and Philosophy* 13, 137–177.
- Huang Y (1994). *The syntax and pragmatics of anaphora*. *Cambridge Studies in Linguistics*. Cambridge: Cambridge University Press.
- Huang Y (2000). *Anaphora: a cross-linguistic study*. *Oxford Studies in Typology and Linguistic Theory*. Oxford: Oxford University Press.
- Kamp H & Reyle U (1993). *From discourse to logic: introduction to model theoretic semantics of natural languages, formal logic and discourse representation theory*. Dordrecht: Kluwer.
- Karttunen L (1976). 'Discourse referents.' In McCawley J (ed.) *Syntax and semantics 7: notes from the linguistic underground*. London: Academic Press. 363–385.
- Kempson R (1988). 'Logical form: the grammar cognition interface.' *Journal of Linguistics* 24, 393–431.

Counterfactuals

S Barker, University of Nottingham, Nottingham, UK

© 2006 Elsevier Ltd. All rights reserved.

Counterfactuals are a class of conditionals, or *if-then* statements. They are interesting because they are instances of a semantically significant class of sentences, conditionals, and because of their intimate connections to modality, laws of nature, causation, dispositions, knowledge, perception, and other concepts, all of which are of central philosophical concern.

Counterfactuals are those *if*-sentences that have a modal auxiliary as main consequent verb and have in their antecedent and consequent clauses a backward shift of syntactic tense relative to notional tense that renders those clauses incapable of self-standing assertion (see Dudman, 1994). Some instances:

- (1) If Osama had not existed, George would/might/could have created him.

- (2) If Osama were to strike, George would defeat him.

Paradigmatically, counterfactuals are uttered in the knowledge that they have false antecedents, but this is not a characterizing mark. They may be issued when their antecedents are believed to be merely improbable. Furthermore, *indicative* conditionals may be asserted with known false antecedents; consider (3).

- (3) If Bill Clinton was bald, no one knew about it.

There is some debate about where to draw the line between counterfactuals and, more broadly, subjunctives and indicative conditionals (see Jackson, 1987; Dudman, 1994; Bennett, 2003). This mainly concerns the status of future open conditionals like (4), which, as Dudman (1994) pointed out, has the same syntactic tense shift found in counterfactuals and has a consequent modal auxiliary.

- (4) If Clinton goes bald, everyone will know about it.

Another issue is whether counterfactuals have truth conditions. There are strong arguments that indicatives, like (3), do not but that they do require a probabilistic assertability condition semantics (see Adams, 1975; Edgington, 1995). Unification of *if* indicates that counterfactuals should be treated similarly. But just how to provide a probabilistic semantics for counterfactuals is far from obvious (see Edgington, 1995; Barker, 1999).

Let us assume from now on that counterfactuals do have truth conditions. The central problem of counterfactuals is devising a noncircular specification of their truth conditions. Philosophers have tended to concentrate on the *would*-conditionals, conditions of the form *If P had been the case, Q would have been* (abbreviated $[P > Q]$), the assumption being that other counterfactuals, such as *might*-conditionals (*If P had been the case, Q might have been*, abbreviated $[P \Diamond \rightarrow Q]$), can be analyzed in terms of the former (for proposals, see Lewis, 1973, 1986c; Stalnaker, 1981).

There are broadly two forms of analysis of counterfactual truth conditions: *metalinguistic* analyses and *possible worlds* analyses. Both are said to be inspired by the idea, in Ramsey (1990 [1929]), that we evaluate a conditional by adding the antecedent to our stock of beliefs and adjusting for consistency, to determine if the consequent is in the resultant state.

The older approach is the metalinguistic one. Possible worlds are currently the entrenched view. Lewis (1973) argued that they are equivalent, which is open to dispute (see Barker, 1999).

Metalinguistic Approaches

On the metalinguistic approach, counterfactuals are like condensed arguments – roughly, P and laws of nature L plus facts cotenable with P (legitimate factual premises) entail, or probabilize, Q :

$$(P > Q) \text{ is true} \equiv \{P + \text{Cotenable factual premises} + L\} \rightarrow (\text{probable}) Q$$

The challenge for metalinguistic theories is (i) to define in a noncircular way the conditions for A 's being cotenable with P and (ii) to fashion a conception of law. Goodman (1965) famously found both of these to be insuperable challenges.

Goodman argued that the extensionally correct specification for a premise A to be cotenable with P is as follows:

$$A \text{ is cotenable with } P \equiv (P > \text{still } A)$$

The conditional $(P > \text{still } A)$ is a *semifactual*: a counterfactual with a true consequent A , expressing that

P is causally benign with respect to A , as in *If the match had been struck it would still have been dry*. Goodman argued that the truth conditions of a semifactual are as follows:

$$(P > \text{still } A) \equiv \sim(P > \sim A) \ \& \ A.$$

But this introduces circularity into the analysis, since we have defined conditions of cotenability in terms of counterfactuals.

The second challenge – finding some acceptable analysis of natural laws – is to provide an explanation of why some generalities support counterfactuals and others do not. Goodman despaired of finding that mysterious counterfactual supporting feature. The current conception is that the problem of law is less acute, since irreducible modalities are now more palatable (for discussion of some of the issues, see Jackson and Pargetter, 1980; Sober, 1988).

After Goodman, various metalinguistic theories attempted to deal with the problem of cotenability, e.g., Pollock (1981), but were not successful. Kwart (1986) represented a very sophisticated breakthrough. Central to his analysis was the idea that counterfactuals presuppose indeterminism for their literal truth, since they involve a conception of reality's branching off from a time t_0 prior to the antecedent time t_p and developing through a *scenario of transition* to P – and orderly, lawful development of reality to P . Kwart showed that Goodman's cotenability condition was extensionally incorrect, since his truth conditions for semifactuals were wrong and he analyzed cotenability by using the notions of positive causal relevance and causal irrelevance, which in turn were reduced to probabilistic relations. Kwart's theory, however, could not capture certain cases (see Belzer, 1993).

Barker (1999) provided an improvement over Kwart with a quasialgorithmic approach to solving the problem of cotenability; Barker's approach invoked causal or more broadly connective relations. The truth conditions of a semifactual relative to a scenario S leading to P are given thus, where p_i is an event in S :

$$(P > \text{still } A) \equiv \sim(P \Diamond \rightarrow p_i \text{ causes } \sim A) \ \& \ A$$

To determine $(P > Q)$ relative to a scenario S , we need to find an instantiated law-based generality ($G1$) linking P to Q via true A . To determine A 's cotenability, we need to evaluate $(P > \text{still } A)$. To determine that, we need to determine $(P \Diamond \rightarrow p_i \text{ causes } \sim A)$. This can be evaluated directly if there is no instantiated law-based causal generality ($G2$) linking P , factual premise B , and the possibility of p_i causing $\sim A$. If there is no ($G2$), $(P \Diamond \rightarrow p_i \text{ causes } \sim A)$ is false, $(P > \text{still } A)$ is

true, and so $(P > Q)$ is true. If there is a (G2), then we ask if the might-conditional cotenability condition $(P \Diamond \rightarrow \text{still } B)$ is true, and that leads to a line of inquiry similar to that for $(P > \text{still } A)$. The recursion is bound to terminate at some stage and provide a determinate answer about the truth values relative to S of the counterfactuals in the procedure. We do the same evaluation of $(P > Q)$ for all the scenarios leading to P . A noncircular determination of $(P > Q)$'s true results. Unlike Kvart's account, Barker's approach can be extended to deal with probabilistic counterfactuals.

Possible Worlds Approach

Lewis (1973) proposed these truth conditions in terms of possible worlds and similarity relations:

$(P > Q)$ is true iff some (accessible) world where both P and Q is more similar to our actual world @, overall, than is any world where P is true and Q is false.

Stalnaker (1968) offered a variant that assumes that there is a unique closest P -world to the actual: Lewis's does not.

Lewis (1973) gave some evidence that this general approach captures the basic logic of counterfactuals – explaining the failure of transitivity, antecedent strengthening, etc. (Metalinguistic theories can do similarly through relativizing entailments to scenarios of transition.)

According to Lewis, would-counterfactuals have two readings: *forwardtrackers* and *backtrackers*. Forwardtrackers are meant to be those counterfactuals that are used in analyzing causation and that capture the sense in which the future depends causally on the past. Backtrackers carry information about the causal in the opposite direction.

Lewis (1986c) offered R as the criteria governing the similarity metric that determines relative similarity of possible worlds in the case of forwardtracking counterfactuals. R :

- i. It is of first importance to avoid big, widespread, diverse violations of law.
- ii. It is of second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
- iii. It is of third importance to avoid even small, localized, simple violations of law.
- iv. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

The result of application of R is meant to be that the closest P -worlds to the actual @ will be worlds

that diverge from @ at a time t_0 not too long before t_p , at which point a small miracle of divergence occurs, and thereafter develop in a law-governed way. *Miracle* has purely a technical meaning: miracle relative to the laws of @, not with respect to the laws of the P -worlds, which remain unbroken. Laws of nature require nothing stronger than a Humean account. Laws are simply statements that feature in the most economical description of the pattern of matters of fact in a world in which perfectly natural properties are picked out.

Unfortunately, there is strong evidence that Lewis's analysis fails for deterministic cases (see Elga, 2000; Hausman, 1998), because the *de facto* temporal asymmetries of physical determination that Lewis thought provide the basis for counterfactual dependency of future on past, do not in fact deliver that result. There are strong reasons to think it fails for indeterministic cases and in particular for counterfactuals such as (5), which depends for its truth on a chance outcome: the coin's landing heads.

(5) If I had bet on heads I would have won.

If, following clause (iv) of R , approximate agreement of fact after t_p counts for nothing, Lewis's theory deems (5) false, because (I bet heads)-worlds in which the coin lands heads and I win and (I bet heads)-worlds where it lands tails and I lose are equally similar to the actual world. If we allow approximate agreement of fact to count, then it is still not obvious why (5) would come out true. Change the situation so that the end of the world will occur if I win, then (5) comes out false, even though my betting has no causal influence on the coin. It seems global similarity has absolutely nothing to do with the evaluation of (5). Indeed, it can be argued that causation is essential (see Barker, 2003). Lewis's analysis fails.

Some Issues

A pure analysis of counterfactuals without invoking causation looks dubious. There is real concern that Lewis's (1986b) project of a counterfactual analysis of causation will be circular (see Barker, 2003). A more general issue is the direction of counterfactual dependency. To date, both metalinguistic and possible worlds treatments have failed to provide an adequate account. Why does counterfactual thought involve divergence from past to future rather than the other way around? (For one interesting discussion of these matters, see Price, 1996.)

See also: Conditionals; Possible Worlds: Philosophical Theories.

Bibliography

- Adams E (1975). *The logic of conditionals*. Dordrecht: Reidel.
- Barker S (1999). 'Counterfactuals, probabilistic counterfactuals and causation.' *Mind* 108, 427–469.
- Barker S (2003). 'A dilemma for the counterfactual analysis of causation.' *Australian Journal of Philosophy* 81, 62–77.
- Belzer M (1993). 'Kvart's a theory of counterfactuals.' *Nous* 27, 113–118.
- Bennett J (2003). *A philosophical guide to conditionals*. Oxford: Oxford University Press.
- Dudman V H (1994). 'On conditionals.' *Journal of Philosophy* 91, 113–128.
- Edgington D (1995). 'On conditionals.' *Mind* 104, 235–329.
- Elga A (2000). 'Statistical mechanics and the asymmetry of counterfactual dependence.' *Philosophy of Science, suppl. vol. 68*, 313–324.
- Goodman N (1965). *Fact, fiction and forecast*. Cambridge, MA: Harvard.
- Hausman D M (1998). *Causal asymmetries*. Cambridge: Cambridge University Press.
- Jackson F (1987). *Conditionals*. Oxford: Basil Blackwell.
- Jackson F & Pargetter R (1980). 'Confirmation and the nomological.' *Canadian Journal of Philosophy* 10, 415–428.
- Kvart I (1986). *A theory of counterfactuals*. Indianapolis: Hackett Publishing.
- Pollock J (1981). 'A refined theory of counterfactuals.' *Philosophical Studies* 74, 239–266.
- Lewis D (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lewis D (1986a). *Philosophical papers* (vol. 2). Oxford: Oxford University Press.
- Lewis D (1986b). 'Causation.' In Lewis (ed.) (1986a). 159–213.
- Lewis D (1986c). 'Counterfactual dependence and time's arrow.' In Lewis (ed.) (1986a). 32–66.
- Price H (1996). *Time's arrow and Archimedes' point: new directions for the physics of time*. New York and Oxford: Oxford University Press.
- Ramsey F (1990 [1929]). 'General propositions and causality.' In Mellor D H (ed.) *Philosophical papers*. Cambridge: Cambridge University Press. 145–163.
- Sober E (1988). 'Confirmation and lawlikeness.' *Philosophical Review* 97, 93–98.
- Stalnaker R (1968). 'A theory of conditionals.' In *Studies in logical theory, American philosophical quarterly*. Monograph 2. Oxford: Blackwell. 98–112.
- Stalnaker R (1981). 'A defence of conditional excluded middle.' In Harper W *et al.* (eds.) *Ifs: conditionals, belief, decision, chance and time*. Dordrecht: Reidel. 87–104.

Creativity in Language

E V Clark, Stanford University, Stanford, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

Speakers change languages all the time because their needs in using language change. One major type of change appears in the words people use and the meanings they assign to them. Speakers construct new words to add to their vocabulary, to talk about new trends and developments in society (*boycott*, *sabotage*, *space-station*) and to take into account technical and scientific innovations (*aviation*, *gene-splicing*, *parallel processing*, *flash-memory*). And they discard terms that no longer serve any purpose – some that are rarely used today include *barouche*, *jeroboam*, *wringer*. This is why dictionary making is endless. Lexicographers track how speakers change their words over time, and the vocabulary of a language is never fixed. Speakers continually add new terms for new categories and adjust existing meanings to take account of extensions (often metaphorical) and also shifts introduced by the addition of other words in the same domains, for example the shift from *guitar* to

acoustic guitar under pressure from the contrast with *electric guitar*. Some new terms that speakers coin are taken up by others, eventually by the speech community as a whole, and so become part of the current conventional vocabulary. Other coinages are used once only, and then vanish again. In all of this, speakers effectively assign an indefinite number of meanings to a finite number of forms in the language they use. They do this by combining compositional meanings with pragmatic inferences dependent on the context, and so continually extend the resources available in new ways.

Coining Words

When speakers coin words, they do so to convey some meaning that is not covered by the meanings of existing words in the vocabulary. For example, a fairly new verb like *to farmer*, meaning 'play the role of farmer (without being a real farmer)' contrasts with its near-neighbor *to farm*, which designates the activity of being a (real) farmer. In constructing the forms of new words, speakers follow the rules

for word formation in their language. In English, for example, they can have recourse to novel compound nouns (e.g., *the circle-group* vs. *the oval-group*, to describe two groups distinguished by insignia), to derived nouns and adjectives (e.g., among recent nouns *a peace-keeper*, *the polluter*; among recent adjectives *internationally minded*, *safety-challenged*), and to denominal verbs (*to medal* in an Olympic event, *to land* a plane even on water, *to e-mail* the message) (e.g., Clark and Clark, 1979; Marchand, 1969). Other languages may draw on a slightly different repertoire of options: French, for example, like other Romance languages and like Hebrew, makes little use of compounding and relies much more extensively on derivation for new words (e.g., Depecker, 2001; Guilbert, 1975). In making use of such word-formation patterns, speakers generally choose the most productive forms – the patterns currently favored by speakers – for the meanings to be expressed (Clark, 1993). This productivity of a pattern can change over time. For example, the Verb + Noun pattern used for forms like *pinch-penny*, *toss-pot*, and *pick-pocket* lost its productivity after this pattern acquired a generally negative connotation in the 1700s. Other patterns, such as N + V-er for agentive and instrumental meanings (as in *tomato-grower*, *cheese-cutter*, *picture-painter*), may remain highly productive over many years. But speaker preferences for particular patterns and affixes can shift, and this typically results in changes in the productivity of particular word-formation patterns and suffixes over time (Dubois, 1962).

Coinages are ubiquitous in everyday speech and in many kinds of writing. While temporary or nonce uses come and go, other coinages may be regulated in that language academies often make recommendations about new terminology. This happens in many countries for new fields of enquiry where a language academy may simply keep track of the coinages that speakers come up with as the Swedish Academy does, or make active recommendations about the terms to use in specific domains, as in France or Israel (Berman, 1997; Clark, 1993). Most language academies tend to focus more on written than on spoken forms of the target language, yet it is in their everyday speech that people continually renew their language.

Children Coin Words, Too

Children start learning what the options are for constructing words – the ways to put together stems and affixes to express different meanings – from the earliest stages of acquisition. They may make use of some productive options in their language from as young as age 2. Like adults, they coin words

to add to their vocabulary and so extend their range in talking about the objects and events around them. In doing this, they apply patterns observable in the established vocabulary in constructing new words: nouns such as *sleepers* for ‘pyjamas’, *climber* for ‘ladder’, *reacher* for ‘someone who can reach a long way’, and *drummer* for ‘drummer’, are typical of the forms constructed by 2- and 3-year-olds. They also construct new adjectives such as *windy* for ‘blown by the wind’, or *walky* for a path ‘one can walk along’, as well as new verbs like *crack out* for ‘hatch’, *oar* for ‘row’, *piano* for ‘play the piano’, or *untight* for ‘loosen’.

Young children have much more limited vocabularies than adults at first, so they may coin many words illegitimately. That is, they coin a word to convey a meaning for which there is already a conventional term. They may come up with forms like *farsee-er* for ‘telescope’, *to car* for ‘to drive’, *to rug* for ‘to vacuum [a rug]’, or *to needle* for ‘to mend’ (Clark, 1993). These coinages are pre-empted by the conventional terms adults use for those meanings, and children give them up as they learn the appropriate terms. As children get older, they become more adept at coining words where there are no existing words with just the meanings they wish to express. They become sensitive to the relative productivity of specific patterns and affixes, and, like adults, favor more productive over less productive options. Also like adults, they continue to attend to transparency in meaning – 2-year-olds prefer a compound like *magic-man* over a derived form like *magician* in which the term *magic* is less discernible. They also attend to simplicity in form – 2-year-olds prefer making the least possible change when they construct a new word form.

Language Revival

The 20th century saw the revival of several languages. The best known of these perhaps is Modern Hebrew, adopted by early Jewish settlers in the Near East as their everyday language from the 1890s on. After the establishment of the state of Israel in 1948, Modern Hebrew was adopted as the national language. Speakers of Hebrew then had to create hundreds of new words for all the everyday things they needed to be able to talk about in modern life, from tractors and engine parts to postal systems and school-leaving exams. They borrowed extensively from other languages, through loan translations, and they coined words for whatever they needed in agriculture (including fish farming, raising turkeys, and meat packing), plumbing, child rearing, schooling, and marketing, as well as in politics, economics,

and science (Blau, 1981; Bolozky, 1999; Ravid, 1995). Israel also set up an Academy of the Hebrew Language to make recommendations about new vocabulary as speakers expanded the existing resources from historical and ritual varieties of Hebrew. The Academy continues to make recommendations today, often lagging behind speakers who take the initiative before any official recommendation appears.

The Irish Gaelic language in Eire faced a similar challenge as the government pushed for its revival as people's everyday language (Watson, 2003; Wright, 1996). Reintroducing Irish in geographic areas where it had not been spoken for several generations, and adding the range of vocabulary needed so it could indeed be used for all everyday purposes presented a formidable challenge. Speakers who had retained their Irish tended to be from rural areas and rural occupations, so, as in Israel, the need for new vocabulary in many social and scientific domains was extensive. Welsh faces similar challenges in maintaining itself as a full everyday language (Aitchison, 2000), as does Scots Gaelic and many other minority languages that often received little official support.

Other languages currently undergoing similar attempts at revival include a number of Australian aboriginal languages and American Indian languages (e.g., Amery, 2000; Hinton and Hale, 2001). In many cases, linguists have been called on to help devise writing systems and help prepare teaching materials, so the languages in question can be (re-)introduced into the community at the nursery school level and up, within the local school system. Few of these revivals, though, have gone as far as Hebrew or Irish Gaelic in building up vocabularies adequate for all facets of modern life. Yet this infusion of vocabulary may be a vital ingredient for endangered languages: speakers' willingness to invigorate the language with new words may well be essential if a language is to remain viable.

Syntactic Units and Combinations

Do speakers construct entirely original, new utterances on each occasion, using syntactic rules to put together new combinations of terms, or do they rely more on ready-made phrases, chunks, and idioms? Within generative grammar, the traditional claim has been that speakers can generate innumerable utterances that they have never produced before (e.g., Chomsky, 1986). That is, speakers are assumed to construct original syntactic combinations, each one built from scratch according to the rules. But does that claim really capture what goes on in most interactions? As native speakers, people have access to a large repertoire of words, phrases, idioms, and

constructions, many of them very frequent in everyday speech.

So an alternative view is that speakers typically build their utterances from ready-made routines and a limited set of constructions. In relying on formulaic uses and routines in language, speakers tailor them to their needs by inserting words, idioms, and phrases that result in new combinations within a construction, and even thereby stretch the meaning of a construction. One reason for doing this is to take advantage of the chunks of language people store in memory (Bolinger, 1976; Fillmore, 1979). This view is consistent with recent work on the myriad syntactic forms, often designated as peripheral to a language, where 'the rules fail' (Culicover, 1999). Recent approaches have tackled these as exemplifying cases where a construction-based, rather than a rule-based, approach may prove more explanatory of how speakers use language. The hypothesis is that when people talk, they work with constructions and formulaic chunks where they can fill variable slots to express specific meanings on different occasions (e.g., Kay and Fillmore, 1999).

Extending Language in New Directions

Creativity in language – stretching meanings and coining words – is also a form of play, one delighted in by many writers who have stretched their language to new limits. In James Joyce's *Ulysses*, Keri Hulme's *The Bone People*, or Tibor Fischer's *The Thought Gang*, for example, each writer extends the meanings of existing words through new juxtapositions and figurative uses. And each writer coins many new words to express new meanings. Poets like Gerard Manley Hopkins (*daylight's dauphin, dapple-dawn-drawn Falcon*, or *golden grove unleaving*) or Dylan Thomas (*the windfall light, the round/ Zion of the water bead/ And the synagogue of the ear of corn*) are often major innovators in language, stretching meanings and adding new forms for new meanings. And many playwrights also extend the uses of a language. Here English speakers probably owe their greatest debt to Shakespeare whose plays are the source of innumerable metaphors, figures of speech, and phrases generally taken for granted until we are startled to hear them fresh-minted in a speech from one of his characters.

The extent to which speakers exploit established and novel figurative uses of language in everyday speech is unclear, but research on language processing shows that readers and listeners readily interpret new uses, figurative uses, and extensions of existing meanings without difficulty (e.g., Gerrig, 1993; Gibbs, 1994).

In summary, speakers are creative in their language use because language is never a precise map of what they wish to convey. Languages do not all select the same elements in a scene to encode, but they select enough to evoke events and their participants. Yet this mapping leaves unmentioned many a detail, as well as many changes in society itself. Speakers, writers, poets, and children remedy such gaps by stretching the meanings of the terms they have and by constructing new words for the new meanings they wish to convey.

Bibliography

- Aitchison J W (2000). *Language, economy, and society*. Cardiff: University of Wales Press.
- Amery R (2000). *Warrabarna Kurna! Reclaiming an Australian language*. Lisse: Zwets & Zeitlinger.
- Berman R A (1997). 'Modern Hebrew.' In Hetzron R (ed.) *The Semitic languages*. London: Routledge. 312–333.
- Blau J (1981). *The renaissance of modern Hebrew and modern standard Arabic*. Berkeley: University of California Press.
- Bolinger D (1976). 'Meaning and memory.' *Forum Linguisticum* 1, 1–14.
- Bolozky S (1999). *Measuring productivity in word-formation*. Leiden: Brill.
- Chomsky A N (1986). *Knowledge of language*. Cambridge: MIT Press.
- Clark E V (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Clark E V & Clark H H (1979). 'When nouns surface as verbs.' *Language* 55, 767–811.
- Culicover P (1999). *Syntactic nuts: hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.
- Depecker L (2001). *L'invention de la langue*. Paris: Colin-Larousse.
- Dubois J (1962). *Etude sur la dérivation suffixale en français moderne et contemporain*. Paris: Larousse.
- Fillmore C J (1979). 'On fluency.' In Fillmore C J *et al.* (eds.) *Individual differences in language ability and language behavior*. New York: Academic Press. 85–110.
- Gerrig R J (1993). *Experiencing narrative worlds*. New Haven: Yale University Press.
- Gibbs R W (1994). *The poetics of mind*. Cambridge: Cambridge University Press.
- Guilbert L (1975). *La créativité lexicale*. Paris: Larousse.
- Hinton L & Hale K (eds.) (2001). *The green book of language revitalization in practice*. New York: Academic Press.
- Kay P & Fillmore C J (1999). 'Grammatical constructions and linguistic generalizations: The What's X doing Y? construction.' *Language* 75, 1–33.
- Marchand H (1969). *The categories and types of present-day English word-formation* (2nd edn.). Munich: Becksvetlag.
- Ravid D (1995). *Language change in child and adult Hebrew*. Oxford: Oxford University Press.
- Watson I (2003). *Broadcasting in Irish: minority language, radio, television, and identity*. Dublin: Four Courts Press.
- Wright S (ed.) (1996). *Language and the state: revitalization and revival in Israel and Eire*. Clevedon: Multilingual Matters.

This page intentionally left blank

D

Data and Evidence

C Schütze, University of California at Los Angeles,
Los Angeles, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

This article surveys the major kinds of data used by linguists, focusing on the properties that bear on their usefulness in linguistic analysis – what they can serve as evidence of, and what sorts of assumptions underlie their interpretation. Linguistic data is divided herein along two dimensions: its origin, i.e., the method or technique by which it was obtained, and its kind, i.e., the phenomenon from which it was drawn. Although these dimensions are in practice not fully orthogonal, the separation helps make the issues clearer.

Sources of Data

Corpora

‘Corpus data’ can be characterized as (a collection of) preexisting samples of language – that is, language that was originally produced for some purpose other than directly answering linguists’ questions.

The contents of corpora provide evidence that has not been biased by the researcher. Similarly, they can provide samples of language produced when the speakers/writers were not consciously reflecting on the form of their utterances. (This point is not trivial: Labov (1975) documented that speakers who denied that they would ever produce a certain construction went on to do so spontaneously during the very same interview.) Corpora may bring to attention phenomena that would not otherwise have come to mind. Because they typically contain extended passages of contiguous language, they supply evidence about phenomena that cannot be studied in isolated words or sentences, e.g., how referents are identified throughout a narrative or discourse. They also provide virtually the only kind of evidence about languages that are no longer spoken.

The ability to search a corpus by computer and calculate statistics on its contents provides detailed quantitative data – for example, about the relative

frequencies of different forms or constructions – that cannot be directly obtained by other methods discussed in this section. For example, corpus data are increasingly being used to show that constructions that tend to sound artificial and awkward when presented as isolated examples in linguistics articles nonetheless may occur and sound natural in everyday language situations. A further benefit to corpora of spontaneous language is reflected particularly in studies of acquisition and speech production: One can find ‘errors’ in such corpora, whether systematic productions by children that are not possible in the adult language or sporadic speech errors (slips of the tongue).

The kind of evidence that a corpus can supply is limited, however: for a given form of interest (morpheme, word, construction, etc.), it can tell us whether or not that form occurs, and if so, how frequently and in what contexts. What such occurrences and nonoccurrences are evidence of has no general answer. Nonoccurrence in particular is a challenge to interpret: it could imply that the form is not possible in the language, that the intended message never called for its use, or that it is inappropriate to the genre/register/modality or some other property of the corpus. Absence from even a large corpus constitutes at best a weak piece of evidence for ill-formedness.

Taking the **presence** of a form in a corpus as evidence that it is part of the language might seem more straightforward, but here too there are caveats, since by definition there is no way to further question the person who produced the utterance to begin with. It is paramount to know the exact origins of the corpus. For example, many texts of dead languages have been copied by people who did not themselves speak the language in question; they might have (knowingly or accidentally) altered the text while copying it. Similarly, when searching the Internet, it is hard to be certain whether the source of a given utterance is a native speaker. Furthermore, to obtain evidence that is relevant to linguistic theories, one typically also needs to know what an attested form means. This may or may not be determinable from the corpus.

For practical reasons, most corpora today are collections of written rather than spoken language. Written language in general admits of ill-formedness that would not occur in speech, arising from the writing and editing process. In the composing or editing of text it is possible to make piecemeal alterations to a sentence without being forced to reread it to verify that it still hangs together as intended (as well as, of course, making typographical errors). Editing may also introduce structures that would not be producible in real time, such as multiple center-embeddings. Thus, a single instance can scarcely count as evidence of anything whatsoever. The typical course of action is to count: if a form of interest occurs enough times in a corpus (there is no standard for how many times is enough), it can be assumed not to be the result of a random editing or copying error; however, the confounds noted above could still apply. (A ‘confound’ is a factor that systematically affects the outcome of a study but is not deliberately manipulated by the researcher. In the present case, the outcome could consist of the distribution of words, sentence types, etc., in a corpus. This could be systematically affected by, for example, the editing process, which may be totally outside the scope of the researcher’s interests if these involve spontaneous language production. What the researcher is manipulating might be anything that varies in spontaneous production, e.g., sentence length or who is speaking.)

Grammaticality Judgments

So-called ‘grammaticality judgments’ involve explicitly asking speakers whether some particular string of words is a possible utterance of their language, with an intended interpretation either implied or explicitly stated. (Analogous judgments can be sought for materials larger or smaller than a single sentence.) Grammaticality judgments provide evidence about the status of phenomena that occur so rarely in spontaneous language use that we could not learn about them from studying corpora; in particular, they distinguish possible from impossible utterances among those that have never been naturally produced. Grammaticality judgments sometimes demonstrate knowledge of language in speakers whose behavior on other tasks does not evince the same degree of knowledge. For example, Linebarger *et al.* (1983) showed this experimentally for people with agrammatic aphasia, and it has been reported (e.g., Berko and Brown, 1960) that children will sometimes reject an adult’s faithful repetition of an utterance they have just produced, apparently realizing that their own productions are not hitting the intended target. Another circumstance where speakers’ judgments may diverge from their own usage, noted

above, arises particularly when the phenomenon under study involves forms identified with social status. A way in which one can try to deal with such cases is to distance the judgment from the speakers themselves, by not asking, “Would you say this?” but rather, “Have you heard people say things like this?” or “Do you know anyone who talks like this?”

Although there are strings of words that seem unequivocally completely well-formed (*Mary saw John*) and strings that seem completely ill-formed (e.g., *Fred the the this*), most strings of linguistic interest lie in a vast gray area between these two extremes. For this reason, simply eliciting yes/no grammaticality judgments is of limited value. Two solutions are to elicit responses on a five- or seven-point scale (as is typical in psychology), or to elicit relative judgments that compare minimal pairs of sentences. The notation used to indicate the status of sentences has evolved to allow numerous gradations of well-formedness to be indicated, but even so, the **relative** status of examples is primarily what is being described; the same annotation does not always convey the same absolute level of well-formedness.

In some cases it is difficult to replicate relatively agreed-upon judgments of linguists while testing naïve subjects with no linguistic training (e.g., Gordon and Hendrick, 1997). Partly for this reason, and partly for sheer convenience, linguists rely heavily on other linguists for judgment data on some languages. Some consider this undesirable (cf. Spencer, 1973). Valian (1982) made a case in favor of using such ‘expert’ judgments, based on analogy to another domain: she noted that wine tasting, for example, relies on the acquired ability to detect subtle distinctions that inexperienced wine drinkers simply cannot make. Wine experts detect these subtleties easily, reliably, and consistently from one person to the next. Her suggestion was that linguists have the same skills with language. One practice that is clearly undesirable, however, is for investigators to use **their own** judgments as their primary evidence.

The nature of grammaticality judgments has sometimes been characterized with reference to traditions in psychology that might lead one to worry about the wisdom of using them as evidence in linguistics: they have been described as “introspective” judgments or “intuitions,” but neither of these terms applies accurately (Carr, 1990; Schütze, 1996). They have more in common with the responses in psychophysics experiments, having the character of sensations or reactions to stimuli, or else reports about mental states (e.g., pain) or social-psychological judgments for which there is no objective external measure to compare them to, e.g., judging the seriousness of various crimes or the prestige of various occupations.

There is also no basis for the occasionally encountered view that grammaticality judgments are somehow related to competence, while spontaneous production and comprehension data are related to performance; all are behaviors – hence performance data – but all can in principle provide evidence (indirectly) about competence (cf. Smith, 2004). For these reasons, Schütze (1996) argued that ‘acceptability reactions’ would be a more accurate and less confusing term, removing reference to the ambiguous and loaded notion ‘grammar,’ but the dominant usage is unlikely to be overturned.

Fieldwork

Much of our linguistic data is vanishing before our very eyes: at least 20% and perhaps as many as 50% of the languages spoken in the world, as of the year 2000, were not being learned by children and will thus cease to be spoken at all by 2100, unless dramatic changes occur (Krauss, 1996). This state of affairs has led to the suggestion that linguists should devote greater attention to documenting and helping to preserve endangered languages, which typically involves fieldwork with native speakers. Field linguists have a great influence over what kinds of evidence can subsequently be constructed from their data, by virtue of how they record the data (e.g., how detailed their transcriptions are), what forms they choose to elicit, and which of these they disseminate. Eliciting meaningful data is a specialized skill and art: one must, for example, avoid inventing utterances in the speaker’s language and simply asking for good/bad judgments about them as one would do with a fellow linguist, because for various sociological reasons speakers tend to over-accept ungrammatical forms. At a minimum, one must verify that speakers can repeat the utterance back; if not, it probably is not really part of the language they speak (Munro, 2000).

Data gathering for ethnographic and sociolinguistic purposes, including dialectology, is also typically considered a type of fieldwork; it mainly involves language as it is naturally spoken when people meet and interact in their daily lives, rather than in a formal elicitation situation. Here, the issues that affect the collection and interpretation of data include a range of variables that are known to influence the way people talk, e.g., the identity of the speaker and addressee (sex, age, social status, education, ethnic heritage, power relationships) and properties of the situation (number of conversants, physical setting, level of formality, topic of conversation, social goals such as establishing group solidarity). To interpret this kind of data, one must face what Labov (1972) called the ‘observer’s paradox’: when people know

they are being observed or recorded, their way of speaking changes, but we want to study how they speak in situations when they are not being observed. A common attempt to minimize this problem involves introducing maximally engaging subject matter to draw speakers’ attention, for example, by asking about emotional topics such as a life-threatening event, or how the world has changed since they were young.

Experiments

It has sometimes been suggested that claims made exclusively on the basis of grammaticality judgment data, do not necessarily bear on how the human language faculty actually works, unless their ‘psychological reality’ has been tested using some other sort of data, typically a formal experiment. This view belies a misunderstanding: grammaticality judgments are themselves data about human behavior that need to be accounted for; they are not intrinsically less informative than, say, reaction-time measures (in fact, one might argue the opposite). The elicitation of grammaticality judgments can itself be seen as a behavioral experiment, albeit one whose generally casual nature may leave it susceptible to certain kinds of confounds. (These can be a result, for example, of the limited number of types and tokens of items tested, the small number of participants questioned, or lack of randomized order of presentation.) Nevertheless, in this section, ‘experiment’ will refer to traditional formal experiments.

The most direct behavioral experiment one can perform to assess the status of sentences or other forms is essentially a more rigorous version of grammaticality judgment collection, typically done using a written questionnaire with a multivalued response scale. In the case of syntactically well-formed strings whose possible meanings are at stake (e.g., sentences with multiple quantifiers), a truth value judgment task may be appropriate for adults, just as it is for children (Crain and McKee, 1985). Beyond experiments that ask participants for their reactions explicitly, a large variety of experimental paradigms used in psychology can be employed to assess participants’ language in more indirect ways. These generally involve a trade-off between reducing the influence of conscious reflection about language on the data and making response measures more difficult to interpret as evidence *vis-à-vis* linguistic issues. There have also been experimental paradigms developed expressly for (psycho)linguistic purposes, e.g., the so-called Wug test (Berko, 1958), lexical decision, self-paced reading, and naming. In some domains, experimental data collection can occur during ‘normal’ language-processing activities (e.g., from eye movements

during reading), where the only ‘abnormality’ involves the physical restrictions imposed by the equipment.

Recent years have shown the promise that we might learn about the workings of language by direct measurements from the brain. Techniques used in this way have included event-related potentials (ERP), functional magnetic resonance imaging, positron emission tomography, and magneto-encephalography. In addition to expensive equipment, these all require special training in the use of the hardware and in statistical techniques for analyzing the data. While some of these techniques may provide a relatively fine-grained characterization of the parts of the brain that are most involved in processing certain aspects of linguistic stimuli, this localization is not terribly helpful as evidence about the particulars of linguistic structure. More provocative have been suggestions (mostly in the ERP literature) that the grammatical status of certain classes of stimuli (e.g., word or sentence types) might be reflected in characteristic patterns of activity (e.g., Neville *et al.*, 1991). For example, it has been claimed that syntactic vs. semantic anomalies have distinct signatures, and even that violations of different grammatical constraints show different patterns. Interpretation of these claims is still open to debate, however, in that it is not established whether a full range of superficially diverse sentence types that by hypothesis share a common grammatical property (e.g., violating Subjacency) have a detectable commonality in brain activity; if it turns out that each kind of ungrammaticality (e.g., extraction from a subject island versus an adjunct island vs. a complex NP island) produces a different ERP pattern, the value of this kind of data for theoretical purposes may be limited. The hope driving this line of research, however, is that if we can find reliable brain signatures for clear cases of certain linguistic properties, we can then compare them to brain measures in reaction to unclear cases and take the brain data as evidence on how to treat those unclear cases.

Kinds of Data

Language Acquisition and Creolization

In addition to informing the study of language acquisition in its own right, data from children acquiring language may be relevant more generally as evidence about human language. By elucidating the ‘starting state’ of the language-acquisition process, it may yield insights into language universals and unmarked values of parameters, on the assumption that the initial state embodies these. Via the course of development, child

grammars may arise that never surface as adult grammars, providing the only opportunity to study those grammars directly. In a similar but more dramatic vein, creolization provides evidence for the noninput-driven nature of knowledge of language, since in this case it is ‘invented’ (unconsciously) by children when there is no robust language present in the input. If one takes the view that language is a mental construct internal to the individual speaker–hearer, then in both normal acquisition and creolization situations, learners are doing the same thing: determining which of their possible internal grammars is most compatible with the ambient speech. What is unique to the emergence of a creole is the fact that the ambient speech is not being generated by any other human’s internal grammar. As a result it might be compatible with a great many of the learners’ possible grammars, or with none at all; in either case the learners must make ‘decisions’ on some basis other than what they are hearing. Thus, even more strongly than for other acquisition data, it is thought that creoles may provide fairly transparent evidence for the defaults of grammar.

Second Language

The language of nonnative speakers may illuminate how native language learning depends on special properties of childhood that do not persist over a lifetime, e.g., so-called ‘critical period’ effects. It may also inform us about which properties of a language should be identified with a single locus of knowledge, e.g., a parameter. As a hypothetical example, consider speakers of languages that have V-final clauses and P-final adpositional phrases. We could pursue the question of whether a single ‘head parameter’ is responsible for ordering in all lexical projections by asking whether such speakers make the same sorts of word-order errors with V as with P when learning a language that is V-initial and P-initial. Adult second-language learners also potentially provide data about language acquisition that are not obtainable from child acquirers due to independent limitations on cognitive ability, memory, attention span, articulatory control of the vocal tract, and so on. Of course it remains to be shown to what extent adult and child acquisition are parallel; in particular, it is hotly debated whether the starting state for second-language acquisition is the universal default from which infants begin, the grammar of the first language, or something else.

Bilingualism

Bilingual (or multilingual) speakers can provide a unique kind of data for linguists, in that they allow

crosslinguistic comparisons that control for all factors other than language (cognitive abilities, cultural experiences, etc.). (This is true also of concurrent bilingual acquisition by children, given the mounting evidence that children separate the systems of their two languages early and easily.) Bilingual speakers additionally have a unique behavior, (intrasentential) code switching, that is constrained in ways that could bear on the grammar. Furthermore, given sufficient understanding of the ways in which two linguistic systems connect with each other and with a (presumably) common cognitive system, new experimental avenues may be opened up.

Language Disorders

Linguistic data from people with language disorders is often (but not uncontroversially) taken as evidence of how language works in unaffected speakers. This may seem paradoxical, since the presence of a disorder *ipso facto* indicates some divergence from the way language normally works. Nevertheless, for many disorders it is believed by numerous researchers that selective loss of language functions reflects the way language is implemented in the brain for affected and unaffected speakers alike, and hence can tell us which aspects of language depend on the same subsystems and which are relatively independent. This assumption might seem justified in the case of brain injuries, as opposed to congenital disorders, but even then, the brain can adapt to such changes, so that a person's speech may actually reflect the loss of some function **plus** an atypical mechanism developed in response to that loss. The fact that data from language disorders typically involve small numbers of speakers, often single case studies, provides further cause for using them cautiously as evidence about human language in general.

Performance Errors

Speech errors have proven to be a rich source of data on the organization of linguistic representations in the mind/brain, as well as on the time course of language production. Certain kinds of speech errors can be experimentally induced, but these are limited and their ecological validity is questionable. Most of the work in this area has traditionally been based on observational collection of spontaneously produced slips assembled into sizeable databases. More recently there has been growing concern about the degree of bias in data collected this way – linguists tend to write down errors that have interesting properties, while letting relatively mundane ones pass by unrecorded. This is widely believed to lead, at the least, to a bias in the relative proportions of error types represented in informally gathered corpora. As a consequence,

some researchers advocate culling errors only from recorded language samples that can be scrutinized exhaustively until no more errors can be found. What this kind of data gains in quality it loses in quantity: the amount of time required for this procedure would have been prohibitive for amassing the collections of tens of thousands of speech errors that were needed to establish basic generalizations. Those errors were produced, and certainly the kinds that occurred in reasonable numbers should be taken as evidence about human language.

Typology and Historical Change

Much can be learned from the range of attested variation across languages, as well as systematically unattested patterns. In addition to providing evidence for the range of possible human languages, such data provide evidence as to which properties of language necessarily covary and which are independent. For this purpose both 'macroparametric' variation (large-scale differences among typologically very different languages) and 'microparametric' variation (minimal differences between linguistically, and usually also geographically, very close languages) is relevant. The kinds of changes that occur in the same language over time can be seen as a special case of the latter, with the further benefit that historical records may allow us to identify the triggers for certain changes, providing evidence for causal connections between language input and grammatical development.

General Remarks

Variability

A problem linguists face every day is how to explain apparently conflicting data about a language, most commonly situations where different speakers of 'the same language' or even 'the same dialect' report different reactions (e.g., different well-formedness judgments) to identical stimuli. On the one hand, those who take a cognitive perspective are investigating what is in the mind/brain of each individual ('I-language'), rather than an abstract object out in the world ('E-language'), so there would be no contradiction in finding that speakers do not have identical I-languages. This could happen because the language input that speakers were exposed to during acquisition was not identical, or because their brains are not identical, i.e., universal grammar is subject to individual variation.

On the other hand, it is worrying if too many such differences are found, because it might be a sign that the judgments are reflecting, in heterogeneous ways, individual differences **outside** the

grammar. The conundrum arises because any data collected from a single individual are susceptible to random confounds; the best way to eliminate such confounds is to consider multiple speakers and hope for convergence. An important emerging finding stressed by Cowart (1997) is that while the well-formedness of a particular sentence might be rated (say) 5 out of 7 points by one participant and 2 out of 7 points by another, there is much less individual variation in the **patterns** of responses – that is, in how sentences rate **relative** to one another. (Cowart also stressed the need to run appropriate statistical tests to determine this.) This is the reason why linguists are primarily concerned with contrasts, say between a sentence marked ‘??’ (barely grammatical/marginally possible) and one marked ‘*’ (ungrammatical/impossible), but not about how good or bad ‘??’ really is in any absolute sense.

Thus, to the extent that numerous speakers show the same pattern of data, the odds of this pattern being the result of some random source diminish rapidly, and it can be taken as solid evidence. However, there is no way to apply the opposite logic: unique judgments of a particular speaker could be good evidence about their idiolect, but it is harder to be certain of this; certainty increases to the extent that the same speaker behaves consistently on different occasions and tasks. More generally, because no single kind of data is perfect, an efficacious approach to linguistic investigation is to seek converging evidence from a wide array of types of data whenever possible.

Data, Evidence, and Theory

There are an essentially limitless number of facts about languages that one could, in principle, document. Data collection therefore always involves selecting – either implicitly or explicitly – certain facts to use as data while ignoring other facts. Often, linguists collect data because they have a theoretical issue they want to pursue and are seeking evidence that will bear on it. To take a syntactic example, linguists studying a particular language might want to know whether a construction they have identified involves (in Government-Binding terms) ‘object control’ versus ‘raising-to-object’ (e.g., *I ordered them to find the answer* versus *I believed them to know the answer*). They are seeking evidence that the crucial noun phrase (*them*) originates as an argument of the ‘upstairs’ verb or evidence that it originates in the ‘downstairs’ clause. They will therefore want to look for data (which might come from grammaticality judgments, if a speaker is available, or from a corpus) on such things as the possibility of the crucial noun phrase being an expletive, an entailment

relation with related sentences lacking the complement clause, synonymy under downstairs passivization, etc. The choice of data to be collected is thus driven by the theory, which in turn determines the questions that are to be answered on the basis of empirical evidence. Even when linguists are not explicitly pursuing any theory, their choices of data to collect are guided by hypotheses or hunches about how languages tend to work, which aspects are relatively unpredictable, and so on, as well as possibly the immediate pragmatic goals of the task. So, if instead the aforementioned linguists had been writing the first grammar of this language, collecting these particular pieces of data might not have occurred to them at all, or if it had, it might have been considered of too little importance relative to other things that such a grammar should include.

A final point to be made about the relationship between data in its raw form and its application as evidence can be illustrated with a well-known, though often misrepresented, example from the early years of generative grammar: the ‘derivational theory of complexity’ (DTC). The idea was that a transformational model of syntax would be independently supported if one could find evidence from language processing for the application of transformations, assuming that a sentence should take longer to process if deriving it required applying more transformations. Data on this point were fairly easy to collect, and in some cases *prima facie* contradicted the predictions of the DTC: for example, passives without *by*-phrases took less time to process than their longer counterparts with *by*-phrases, while the former were supposed to be derived from the latter by one additional deletion transformation; the (full) passive versions of some sentences were sometimes faster to process than their active counterparts – e.g., when the active structure involved center-embedding and the passive converted this to right-branching – while passives were always supposed to be derived from actives by one additional transformation. At the time, these data were taken as evidence against the DTC, which was generally abandoned for several decades; it was concluded that transformations did not necessarily literally apply during the processing of a sentence, either because transformations were the wrong theory of syntax or because syntactic theory as a competence model could not be expected to have reflections in real-time language behavior.

But the leap from the data to what it was evidence of was a mistake. The error involved overlooking a *ceteris paribus* assumption: more transformations should imply more processing time **all else being equal**, but in the crucial cases just mentioned, there is good reason to think that all else was not equal.

For example, processing time surely has a component that increases with the number of words in a sentence (just keep adding instances of the word *very*, for example), which could explain the *by*-phrase result. And center-embedding plausibly puts a burden on short-term verbal memory that could slow language processing in general, independent of any other properties of a sentence, which could explain the passives that were faster than actives. Thus, while the collected data are evidence about the overall processing load differences between pairs of transformationally related sentences, the hypothesis being tested concerned just processing load due to the application of the extra transformation. It therefore called for different data: the processing load differences between sentences that differed in the number of transformations applied, but **not** on any other property that might reasonably be expected to affect processing load. Only the latter kind of data should count as evidence bearing on the DTC.

See also: E-Language versus I-Language; Transformational Grammar: Evolution.

Bibliography

- Berko J (1958). 'The child's learning of English morphology.' *Word* 14, 150–177.
- Berko J & Brown R W (1960). 'Psycholinguistic research methods.' In Mussen P H (ed.) *Handbook of research methods in child development*. New York: Wiley. 517–557.
- Berwick R C & Weinberg A S (1984). *The grammatical basis of linguistic performance: language use and acquisition*. Cambridge: MIT Press.
- Birdsong D (1989). *Metalinguistic performance and interlinguistic competence*. New York: Springer.
- Botha R P (1981). *The conduct of linguistic inquiry: a systematic introduction to the methodology of generative grammar*. The Hague: Mouton.
- Carden G & Dieterich T (1981). 'Introspection, observation, and experiment: an example where experiment pays off.' In Asquith P D & Giere R N (eds.) *PSA 1980: Proceedings of the 1980 Biennial Meeting of the Philosophy of Science Association 2: Symposia*. East Lansing, MI: Philosophy of Science Association. 583–597.
- Carr P (1990). *Linguistic realities: an autonomist meta-theory for the generative enterprise*. Cambridge: Cambridge University Press.
- Chomsky N (1961). 'Some methodological remarks on generative grammar.' *Word* 17, 219–239.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Cowart W (1997). *Experimental syntax: applying objective methods to sentence judgments*. Thousand Oaks CA: Sage.
- Crain S & McKee C (1985). 'Acquisition of structural restrictions on anaphora.' *Proceedings of NELS 16*, 94–110.
- Fillmore C J, Kempler D & Wang W S-Y (eds.) (1979). *Individual differences in language ability and language behavior*. New York: Academic Press.
- Gerken L & Bever T G (1986). 'Linguistic intuitions are the result of interactions between perceptual processes and linguistic universals.' *Cognitive Science* 10, 457–476.
- Gordon P C & Hendrick R (1997). 'Intuitive knowledge of linguistic co-reference.' *Cognition* 62, 325–370.
- Greenbaum S (1988). *Good English and the grammarian*. London: Longman.
- Krauss M (1996). 'Linguistics and biology: threatened linguistic and biological diversity compared.' In McNair L, Singer K, Dobrin L M & AuCoin M M (eds.) *CLS 32: papers from the parasession on theory and data in linguistics*. Chicago: Chicago Linguistic Society. 69–75.
- Labov W (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov W (1975). *What is a linguistic fact?* Lisse, Netherlands: Peter de Ridder [Also published in 1975 as 'Empirical foundations of linguistic theory.' In Austerlitz R (ed.) *The scope of American linguistics*. Lisse, Netherlands: Peter de Ridder. 77–133].
- Levelt W J M (1974). *Formal grammars in linguistics and psycholinguistics* (3 vols). The Hague: Mouton.
- Linebarger M C, Schwartz M F & Saffran E M (1983). 'Sensitivity to grammatical structure in so-called agrammatic aphasics.' *Cognition* 13, 361–392.
- Munro P (2000). 'Field linguistics.' In Aronoff M & Rees-Miller J (eds.) *The handbook of linguistics*. Malden, MA: Blackwell. 130–149.
- Neville H, Nicol J L, Barss A, Forster K I & Garrett M F (1991). 'Syntactically-based sentence processing classes: evidence from event-related brain potentials.' *Journal of Cognitive Neuroscience* 3, 151–165.
- Newmeyer F J (1983). *Grammatical theory, its limits and its possibilities*. Chicago: University of Chicago Press.
- Perry T A (ed.) (1979). *Evidence and argumentation in linguistics*. Berlin: de Gruyter.
- Schütze C T (1996). *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Smith N (2004). *Chomsky: ideas and ideals* (2nd edn.). Cambridge: Cambridge University Press.
- Spencer N J (1973). 'Differences between linguists and non-linguists in intuitions of grammaticality–acceptability.' *Journal of Psycholinguistic Research* 2, 83–98.
- Valian V (1982). 'Psycholinguistic experiment and linguistic intuition.' In Simon T W & Scholes R J (eds.) *Language, mind, and brain*. Hillsdale, NJ: Lawrence Erlbaum. 179–188.

De Dicto versus De Re

E Eaker, University of Western Ontario, London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Philosophers of language and linguists often distinguish between *de re* and *de dicto* readings of certain locutions. For example, take the sentence, ‘Juan believes that the man who wrote *Platform* is Belgian.’ As a rough initial characterization, to read this sentence *de re* is to read it as saying that Juan has a belief *about the man himself* – namely, Houellebecq – that he is Belgian. Juan need not be thinking of Houellebecq as the author of *Platform* in order for the sentence – read *de re* – to be true. Juan may have said, ‘Michel Houellebecq is Belgian,’ or, pointing at Houellebecq, ‘That man is Belgian.’ The important thing is that Juan has a belief about that man, however he is characterized.

In contrast, to read this sentence *de dicto* is to read it as saying that Juan has the following belief: *the man who wrote Platform is Belgian*. Here, ‘the man who wrote *Platform*’ is a crucial part of the content of Juan’s belief. For this *de dicto* report to be true, it is not even necessary that Juan knows who wrote the book *Platform*. Indeed, he may believe that the man who wrote that book is Belgian for reasons that have nothing to do with any information he has obtained about the man Houellebecq.

There are many conflicting ways of characterizing this distinction, and so even a preliminary characterization is controversial. There have been three key elements in traditional discussions of this distinction: syntactic representation, semantic interpretation, and metaphysical import. These three elements are often tied together in such discussions, and it is the aim of this article to provide the reader with tools for keeping them distinct.

Usually, the distinction between *de re* and *de dicto* readings is made for complex sentences, where one sentence, such as ‘Dimitri’s plane is late,’ is embedded in the context of what is often called a sentential-complement-taking clause or a sentential operator. Some sentential operators are ‘Juan believes that...,’ ‘It is not the case that...,’ ‘Howard denies that...,’ and ‘It is possible that...’

Take, for example, the sentence, ‘Juan believes that Dimitri’s plane is late.’ It is commonly thought that there are two readings of this sentence – a *de re* and a *de dicto* reading. On the *de dicto* reading, the embedded sentence is, in a certain important sense, treated as a syntactic and semantic unit. The embedded sentence taken as a whole is treated as a key player in the

semantics of the larger sentence in which it is embedded. To understand the *de dicto* reading then, we must understand what a sentence as a whole can contribute semantically.

Gottlob Frege (1948) proposed that the meaning of a sentence is the thought that it expresses. Bertrand Russell (1905) also was concerned with the thoughts or ‘propositions’ that sentences are used to express. Though Russell talked of propositions as the objects of belief, he thought that they could, in principle, be composed of ordinary objects and properties. In this he differed from Frege, who could make no sense of the idea that an object could be apprehended directly in thought. Frege argued that all thought about objects is mediated by representations of those objects – such representations or senses of expressions occur in Fregean thoughts or propositions, not the objects themselves. So, as key ingredients of a semantic theory, Fregean thoughts and Russellian propositions look like very different sorts of entities. In some writings, Russell seemed closer in his views to those who think sentences express (actual or possible) facts or states of affairs. Contemporary discussions of sentence meaning often involve the question of whether propositions must be ‘structured’ or can be represented as sets of possible worlds (often ‘centered’ with respect to an individual).

In summary, on a *de dicto* reading the embedded sentence is treated as a syntactic and semantic unit. Its semantic value as a whole plays a key role in the semantics of the sentence in which it is embedded – whether this semantic value is taken to be a thought, a proposition, a fact or state of affairs, or an abstract set theoretic entity.

Let us continue with the *de dicto* reading of the sentence, ‘Juan believes that Dimitri’s plane is late.’ If you think, as did Frege, that sentences express thoughts, you will think that this sentence relates Juan to the thought content, *that Dimitri’s plane is late*, and says that Juan believes *that*. If ‘Dimitri’s plane is late’ is embedded in the context of ‘It is not the case that...,’ the embedded sentence might be taken to express the proposition *that Dimitri’s plane is late*. The whole sentence could then be interpreted as the claim that this proposition is not true. If we embed the sentence, ‘Dimitri’s plane is late,’ into the context of ‘It is possible that...,’ this could be interpreted as the claim that the state of affairs represented by the embedded sentence is a possible state of affairs.

These readings, which focus on the embedded sentence and the thought, proposition, fact, or state of affairs it represents, are in contrast with *de re* readings of the same sentences. On a *de re* reading of the sentence, ‘Juan believes that Dimitri’s plane is late,’

Juan is said to believe *of* Dimitri's plane – of that very thing – that it is late. Philosophers, philosophical logicians, and linguists have found various ways of emphasizing the *de re* reading by using different syntactic forms – and often very awkward ones. Applied variously to the sample sentences above, the 'rephrasing' strategy might yield the following: 'Dimitri's plane is believed by Juan to be late,' 'Dimitri's plane is possibly late,' 'Dimitri's plane is such that it is fortunate that it is late,' and 'Howard denies, of Dimitri's plane, that it is late.'

The difference between the *de re* reading and the *de dicto* reading might not strike the reader as important and may be difficult to grasp, even with these reformulations. It becomes apparent, however, when we recognize the distinction between opaque and transparent contexts (or, less theoretically neutral, intensional and extensional contexts.) Consider the sentence, 'Juan believes that Dimitri's plane is late.' Suppose that Pamela Anderson is on Dimitri's plane, though Juan has no idea that this is the case. If so, then Dimitri's plane is also Pamela Anderson's plane. However, would we want to say that Juan believes that Pamela Anderson's plane is late? Contexts that do not allow substitution of co-designative expressions without change in truth value are called opaque. They are also called intensional contexts by those who, inspired by Frege, think that linguistic expressions (including sentences) have both an extension and an intension. According to a Fregean theory, the intension of a sentence is the thought it expresses, and the extension of a sentence is its truth value. What then accounts for the opacity of the sentence, 'Juan believes that Dimitri's plane is late'? That is, why can we not substitute 'Pamela Anderson's plane' for 'Dimitri's plane'? A Fregean would say it is because 'Juan believes that...' is an intensional context – the expressions occurring in this context contribute their intension (thought content) to the semantics of the whole, not their extension (e.g., the plane itself that is being referred to by the two expressions).

However, this is not the whole story for the sentence, 'Juan believes that Pamela Anderson's plane is late.' Russell, for example, recognized a *de re* reading of the sentence. On the *de re* reading, this sentence does not say that Juan believes the proposition or thought content *that Pamela Anderson's plane is late*. It says that Juan believes, of a certain thing, namely a certain plane, that it is late. This plane happens to be Pamela Anderson's plane, Dimitri's plane, and the plane flown by Captain Marshall and First Officer O'Hare. However, on the *de re* reading it does not matter how we refer to the plane, so long as we refer to the right plane. If Juan has a belief about a

thing, then it does not matter how we refer to that thing, so long as we manage to refer to it and to assert correctly what it is that Juan believes about it. Read *de dicto*, 'Juan believes that Dimitri's plane is late,' is opaque, as the substitution of co-designative expressions may not preserve the truth value of the sentence. Read *de re*, it is transparent. Any way of referring to the plane will preserve the truth of the sentence.

Though he used different terminology, Russell (1905) represented the two readings by means of the notion of scope. If the definite description, 'Dimitri's plane,' occurs within the scope of the sentential operator, then it occurs within an opaque context and the whole sentence is to be read *de dicto*. To represent the *de re* reading, Russell put the definite description outside the scope of the sentential operator, where it occurs transparently or extensionally, which means that the substitution of any co-designative (co-extensive) expression will preserve truth. We might gloss the two readings that result from the permutation of the relative scope of the sentential operator and the definite description as follows:

Giving widest scope to the sentential operator to represent the *de dicto* reading:

Juan believes that: Dimitri's plane is late.

Giving widest scope to the description to represent the *de re* reading:

Dimitri's plane is such that: Juan believes that: it is late.

Complex sentences that talk about the beliefs and other psychological attitudes of people are not the only types of sentences that allow for both *de re* and *de dicto* readings. For example, consider the sentential-complement-taking clauses typically called modal operators. Examples of modal operators are 'It is possible that...' and 'It is necessary that...' Consider this sentence, 'It is possible that Dimitri's plane is arriving 15 minutes late.' Read *de dicto*, it is supposed to be understood as the claim that the proposition expressed by the embedded sentence, 'Dimitri's plane is arriving 15 minutes late,' is possible. (When philosophers say that a proposition is possible, they usually mean that though it may not actually be true, it expresses a state of affairs that could have been true or is true in some possible world.) So, even supposing that Dimitri's plane is on time, it could very well have been late – if there was bad weather, congestion on the runway, or the like.

Now, suppose that Dimitri's plane is in fact the only plane that is on time this afternoon. Could we not therefore say, 'It is possible that the only plane that is on time this afternoon is arriving 15 minutes late,' substituting the phrase, 'the only plane that is on time this afternoon,' for the co-designative (co-extensive)

‘Dimitri’s plane’? If the resulting sentence is read *de dicto*, it would seem to claim that the following is a possible state of affairs: The only plane that is on time this afternoon is arriving 15 minutes late. However, this is not a possible state of affairs, because being the only plane that is on time precludes arriving 15 minutes late. If we are reading this sentence *de dicto*, we cannot substitute a co-referring expression for ‘Dimitri’s plane’ without changing what the sentence says and possibly changing its truth value as a result. So, ‘It is possible that...’ seems to be an opaque context too, just as is ‘Juan believes that...’

However, just as ‘Juan believes that...’ allows for a *de re* reading, so does ‘It is possible that...’ If we were to read the modal sentence *de re* – as stating of a certain plane, the very plane that as a matter of fact happens to be the only plane to arrive on time this afternoon – that it is possible that *that plane* should be arriving 15 minutes late, then it does seem to be true. All planes, even planes that happen to arrive on time, are subject to being delayed by 15 minutes. How we refer to the plane does not matter when the claim is a *de re* possibility claim, a claim about what is possible for a particular plane. *De re* modal claims, just as do *de re* claims about Juan’s beliefs, allow truth-preserving substitution of co-designative expressions. This is in contrast to reading the possibility claim *de dicto*, where what is claimed is that a certain proposition or state of affairs is possible. On a *de dicto* reading it matters whether the embedded sentence is ‘the only plane to arrive on time this afternoon is arriving 15 minutes late’ or ‘Dimitri’s plane is arriving 15 minutes late.’ As with belief statements, it has been customary to represent the *de dicto* and *de re* readings of modal sentences by means of Russell’s device of scope:

Giving widest scope to the sentential operator to represent the *de dicto* reading:

It is possible that: the only plane to arrive on time this afternoon is 15 minutes late.

Giving widest scope to the description to represent the *de re* reading:

The only plane to arrive on time this afternoon is such that: it is possible that: it is 15 minutes late.

Notice that one of the sentential operators mentioned in the introduction – ‘It is not the case that...’ – does not create an opaque context. Suppose we read the sentence, ‘It is not the case that John’s uncle is on the plane,’ *de dicto*. That is, we read it as stating that a certain proposition – *that John’s uncle is on the plane* – is not true or as stating that a certain state of affairs does not obtain. Suppose John’s uncle is also Dave’s uncle and Saima’s uncle, and so we substitute ‘Dave’s uncle’ for ‘John’s uncle’ and get this sentence, ‘It is not

the case that Dave’s uncle is on the plane.’ It seems that no matter how we refer to John’s uncle, the truth value of the sentence as a whole is not going to change. If the proposition expressed by ‘John’s uncle is on the plane’ is false, then the proposition expressed by ‘Dave’s uncle is on the plane’ is likewise false; if the state of affairs picked out by ‘John’s uncle is on the plane’ does not obtain, then neither does the state of affairs picked out by ‘Saima’s uncle is on the plane.’ Unlike ‘Juan believes that...’ and ‘It is possible that...’, the sentential operator – ‘It is not the case that...’ – does not seem to create an opaque or non-extensional context.

If that is the case, what is the difference between the *de dicto* reading of the sentence and the *de re* reading in this instance? Suppose we read the original sentence *de re*, as stating of John’s uncle, that it is not the case that he is on the plane. If we are just talking about *that man* (who happens to be John’s uncle, Dave’s uncle, and Saima’s uncle) and saying of him that he is not on the plane, then we can use any way of referring to him that we like. This follows as a matter of course from the *de re* reading. However, we have seen that, though it is forbidden to substitute co-designative expressions in certain *de dicto* contexts, making that substitution here does not affect the truth of what is said. If the difference between the *de re* and *de dicto* readings does not seem to make a difference, then why should we distinguish the two in cases where the operator in question does not seem to create an opaque context?

Here is where it becomes very important to distinguish between the notion of scope – which involves the structural relations among quantifiers, sentential operators, and other linguistic expressions – and the semantic interpretations of complex sentences that scopal disambiguations may or may not be thought to represent. So far we have been using examples where the embedded sentence includes a singular referring expression, such as ‘Dimitri’s plane’ and ‘the only plane to arrive on time this afternoon.’

Yet, consider complex sentences that contain quantifier phrases, such as ‘someone,’ ‘every passenger,’ and ‘most planes.’ The sentence, ‘Juan believes that someone is responsible for the mistake,’ can be read in two ways, usefully disambiguated as follows: (1) ‘There is someone (in particular) whom Juan believes to be responsible for the mistake,’ said when Juan has specifically blamed Geoff; and (2) ‘Juan believes that someone (or other) is responsible for the mistake,’ said when Juan has scoffed at the idea that the mistake was a result of a random computer glitch. The difference between the two readings could be explained as a result of a structural ambiguity concerning whether the quantifier or the operator has widest scope. This is how we represented it above. Alternatively, some have

suggested that the ambiguity should be explained as the result of an ambiguity in the quantifier phrase itself, which might arguably be thought to allow for a ‘particular someone’ or a ‘someone or other’ reading. Likewise consider this sentence, ‘Every robber jumped into a car.’ This also seems to allow for the following two readings, represented by permutations of scope. The first reading is, ‘Every robber is such that there is a car such that that robber jumped into it.’ This would be true if, for example, every robber jumped into a different car. Contrast that reading with, ‘There is a car such that every robber jumped into it,’ which would be true only if every robber jumped into the same car.

In closing, the ambiguities that arise from the interaction of quantifiers with other quantifiers, and from the interaction of quantifiers with sentential-complement-taking phrases, can often be represented by means of the notion of scope (Montague, 1974). In a systematic theory of syntax and the logical structure of sentences, one may wish to acknowledge scopal ambiguities even when the results of scopal disambiguation are logically equivalent. In contrast, the *de re/de dicto* distinction for certain types of ambiguous sentences has typically involved a more robust semantic characterization of the distinct readings that result from disambiguation – even when these distinct readings are represented by means of a structural disambiguation, as above. This can be explained in part historically. The *de re/de dicto* ambiguity has historically been tied to particular sentential operators that have been the focus of philosophical scrutiny because of their close connections with topics central to logic, metaphysics, and epistemology.

The semantic treatment of sentences such as ‘Juan believes that Dimitri’s plane is late’ has been mired in controversy for over 100 years, because of the assumption that the semantics of belief sentences cannot be worked out until a theory of belief is developed. For example, some have argued that *de re* readings of ‘believes that’ sentences do not even make metaphysical sense, on the grounds that there is no direct belief about individuals, but only belief about those individuals under some manner of representing them (Kaplan, 1968). Likewise, the correct semantic treatment of sentences, such as ‘It is possible that Dimitri’s plane is late,’ has been thought to depend on who, among logicians and metaphysicians, has the correct theory of modality (i.e., of necessity and possibility). In fact, some philosophers have argued that *de re* readings of modal sentences do not make metaphysical

sense, because they rely on a purportedly dubious distinction between ‘necessary’ and ‘merely contingent’ properties of objects (Quine, 1956). It is essential therefore, when encountering discussions of the *de re/de dicto* distinction, to be clear on whether the discussion is about syntactic representation, semantic interpretation, or metaphysical import.

See also: Architecture of Grammar; Descriptions, Definite and Indefinite: Philosophical Aspects; Direct Reference; Epistemology and Language; Extensionality and Intensionality; Metaphysics, Substitution *Salva Veritate* and the Slingshot Argument; Object-Dependent Thoughts; Propositional Attitude Ascription: Philosophical Aspects; Referential versus Attributive; Rigid Designation; Scope and Binding: Semantic Aspects.

Bibliography

- Carnap R (1967). *Meaning and necessity: a study in semantics and modal logic*. Chicago: University of Chicago Press.
- Frege G (1948). ‘On sense and reference’ Black M (trans.). *The Philosophical Review* 57, 207–230.
- Geach P (1962). *Reference and generality*. Ithaca, NY: Cornell University Press.
- Kaplan D (1968). ‘Quantifying in.’ *Synthese* 19, 178–214.
- Kaplan D (1971). ‘What is Russell’s theory of descriptions?’ In Pears D F (ed.) *Bertrand Russell: a collection of critical essays*. Garden City, NY: Doubleday Anchor. 227–244.
- Kripke S (1971). ‘Identity and necessity.’ In Munitz M K (ed.) *Identity and individuation*. New York: New York University Press. 135–164.
- Kripke S (1963). ‘Semantical considerations on modal logic.’ *Acta Philosophica Fennica*, 16, 83–94.
- Kripke S (1977). ‘Speaker’s reference and semantic reference.’ In French P, Uehling T & Wettstein H (eds.) *Contemporary perspectives in the philosophy of language*. Minneapolis: University of Minnesota Press. 6–27.
- Montague R (1974). ‘Formal philosophy.’ In Thomason R H (ed.). New Haven, CT: Yale University Press.
- Neale S (1990). *Descriptions*. Cambridge, MA: MIT Press.
- Partee B (1974). ‘Opacity and scope.’ In Munitz M K & Unger P (eds.) *Semantics and philosophy*. New York: NYU Press. 81–101.
- Russell B (1905). ‘On denoting.’ *Mind* 14, 479–493.
- Russell B (1919). *Introduction to mathematical philosophy*. London: George Allen and Unwin.
- Quine W V O (1956). ‘Quantifiers and propositional attitudes.’ *Journal of Philosophy* 53, 177–187.
- Quine W V O (1960). ‘Vagaries of reference.’ In *Word and object*. Cambridge, MA: MIT Press. 138–156.

Default Semantics

K Jaszczolt, University of Cambridge, Cambridge, UK

© 2006 Elsevier Ltd. All rights reserved.

It is hardly contestable that the interpretation of the speaker's utterance by the addressee is frequently driven by the salience of some of the possible interpretations. This salience can be caused by a greater frequency of a certain meaning or by its simplicity, but ultimately it rests on knowledge of social and cultural conventions or the cognitive principles that govern our thinking. Default Semantics concerns such cognitive defaults.

Before laying out the principles of Default Semantics, it is necessary to situate the default-based views in the research on the semantics/pragmatics interface. According to the traditional view, in addition to lexical and syntactic ambiguities, there are also semantic ambiguities such as that between the wide and narrow scope of negation in (1), represented in (1a) and (1b) respectively. 'KoF' stands for 'present king of France'.

- (1) The present king of France is not bald.
- (1a) $\neg \exists x (\text{KoF}(x) \ \& \ \forall y (\text{KoF}(y) \supset y = x) \ \& \ \text{Bald}(x))$
- (1b) $\exists x (\text{KoF}(x) \ \& \ \forall y (\text{KoF}(y) \supset y = x) \ \& \ \neg \text{Bald}(x))$

The ambiguity position, held by Russell, among others, has been successfully refuted. Instead, it has been proposed that such differences in meaning belong to what is implicated rather than what is said (Grice, 1975), and subsequently that semantics can be underspecified as to some aspects of meaning and require pragmatic intrusion in order to arrive at the full propositional representation of the utterance (see, e.g., Carston, 1988, 2002). It is now usual to talk about the *underdetermination* of sense and *underspecification* of the logical form. According to some post-Griceans, such differences in meaning can be explained through default interpretations. The level of defaults has been conceived of in a variety of ways: as belonging (i) to semantics (as in Discourse Representation Theory, Kamp and Reyle, 1993, and its offshoots, such as Segmented Discourse Representation Theory, Asher and Lascarides, 2003) (*see Discourse Representation Theory*); (ii) to pragmatics (Bach, 1994); or (iii) to fully fledged social and cultural conventions, called presumptive meanings or generalized conversational implicatures (Levinson, 2000). All of these default-based approaches advocate some degree of semantic underdetermination, understood as conceptual gaps in the output of lexicon and grammar. In other words, the logical form,

which is the output of the grammatical processing of a sentence, does not provide the totality of meaning of the proposition expressed by the speaker.

While this statement is certainly true, and while it also seems to be true that some pragmatic contribution is often required in order to get the correct truth conditions of the utterance, it does not mean that such an underspecified or underdetermined representation need be distinguished as an epistemologically real level in utterance processing. In Default Semantics, there is no semantic ambiguity, but there is no underspecification either. The logical form as the output of syntactic processing interacts with the information coming from the property of mental states of *having an object*, *being about* something, called their *intentionality*. So, if we ask where meaning comes from, we can point to two sources of meaning: (i) compositionality of the sentence meaning and (ii) intentionality of the mental state that underlies the sentence. Both are equally basic and equally important, and hence it would be incorrect to consider any information coming from intentionality as an additional, pragmatic level of utterance processing. They both belong to semantics. In dynamic approaches to meaning, such as Discourse Representation Theory, such a level of representation, called in Default Semantics an *intentionality-compositionality merger*, has been successfully implemented and seems to be more in the spirit of dynamic meaning than postulating any unnecessary underspecifications or ambiguities (see Jaszczolt, 1999a, 1999b, 2000).

Default Semantics is governed by three main principles: the Parsimony of Levels (PoL), Degrees of Intentions (DI), and the Primary Intention (PI):

- PoL: Levels of senses are not to be multiplied beyond necessity.
- DI: Intentions in communication come in various degrees: they can be stronger or weaker.
- PI: The primary role of intention in communication is to secure the referent of the speaker's utterance.

In PoL, the principle of parsimony with respect to the proposed levels of meaning is taken further than in other post-Gricean approaches. Instead of discerning an underspecified logical form and pragmatic intrusion, both sources of meaning are treated on an equal footing and both contribute to a common level of representation (the intentionality-compositionality merger). DI and PI principles specify how intentionality contributes to the meaning representation. In agreement with the phenomenological tradition (Husserl, 1900–1901), we have defined intentionality as the property of beliefs, thoughts, doubts, etc.,

of being about an object. It is compatible with the definition of intentionality that this aboutness can be stronger or weaker. For example, a definite description 'the best Italian painter' can correspond to a thought about a particular individual, e.g., Michelangelo (and be used *referentially*); to a thought about a particular individual who does not correctly match the description, e.g., Picasso (i.e., there is a referential mistake); or finally to a thought about whoever happens to undergo the description (and be used *descriptively*). In the first case, intentionality is in the strongest form: as a property of the mental state, it reaches, so to speak, a real object. In the middle case, it is weaker: a real object is intended, but there is no such object corresponding to that description, and hence it reaches a mental construct that is a composite of the real person and an incorrect description. In the final case, the intentionality is dispersed and does not reach an object.

Now, intentional mental states need vehicles of meaning, and language is one such vehicle. As a result, linguistic expressions share the property of intentionality, and hence we can talk about intentionality of utterances as well as intentionality of thoughts. On the level of utterances, this intending is realized as *intentions in communication*. Three types of such intentions are distinguished in Default Semantics: an intention to communicate certain content, to inform about certain content, and to refer to objects, states, events, and processes. In accordance with the DI and PI principles, information from the degree of intentionality of the mental state (or the strength of intending, informativeness of an utterance) merges with the information from compositionality and produces the complete propositional representation that conforms to PoL. So, Default Semantics offers a more economical alternative to the approaches founded on underspecified semantics in that it implements Occam's razor (the methodological principle of not multiplying beings beyond necessity) 'one level up.' Semantic representation structures of Discourse Representation Theory have been implemented as formalizations for such intentionality-compositionality mergers (Jaszczolt, 1999b, 2000, in press a).

The DI and PI principles, in recognizing degrees and strengths of intentions, explain how default interpretations can arise. In the case of definite descriptions such as 'the best Italian painter,' the hearer normally assumes that the speaker utters the description with a referential intention and that the description is used correctly. This assumption is further corroborated by the assumed intentionality of the speaker's belief: the intentionality is strongest when a particular, identifiable individual has been intended.

By force of the properties of vehicles of thought discussed in this article, the stronger the intentionality, the stronger the speaker's intentions. In the case of definite descriptions, the stronger the intentionality, the stronger the referential intention. In the case of definite descriptions, there are three degrees of intentionality corresponding to the three readings distinguished previously: (i) the strongest, referential; (ii) the intermediate, referential with a referential mistake; and (iii) the weakest, attributive. The strongest intentionality corresponds to the default reading. This default reading arises instantly, as a compositionality-intentionality merger. Only if addressees have evidence from their knowledge base or from the context that this default is not the case does the default interpretation *fail to arise*. This procedure is an improvement on other default-based approaches where defaults have to be *canceled* or *overridden*. Cancellation of defaults is a costly process and should not be postulated lightly: if there is no evidence of such cancellation, it is better to do without it and assume a more economical model of utterance processing.

Similarly, cognitive defaults can be discerned for belief and other propositional attitude reports. Sentence (2a) can give rise to a report, as in (2b).

- (2a) The best Italian painter painted this picture.
 (2b) Mary believes that the best Italian painter painted this picture.

Using the representation of the Discourse Representation Theory (Kamp and Reyle, 1993; Reyle, 1993), we can represent the possible readings of (2) as in Figure 1 (Jaszczolt, 1999b: 287):

The discourse referent y is enclosed by a box drawn with a broken line, which signals that y can belong to any of the three remaining boxes. If it belongs to the outermost box, the reading is *de re* (about a particular individual, say, Michelangelo). Placed in the middle box, it signals that Mary has a *de re* belief

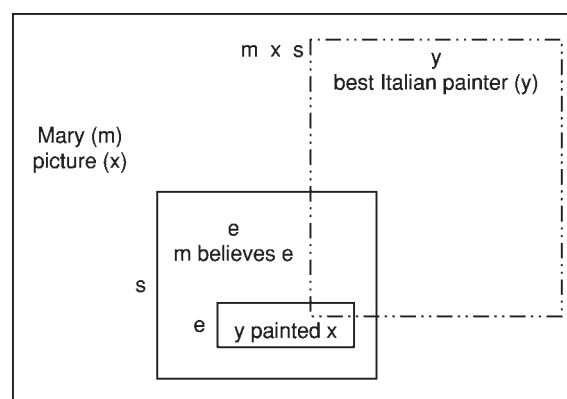


Figure 1 A combined DRS for the three readings of (2b).

but is referentially mistaken, thinking, for example, of Picasso. Placing *y* in the innermost box corresponds to a belief about whoever undergoes the description, i.e., a belief in a proposition (*de dicto*) rather than about a particular individual. Analogously to the case of definite descriptions where referential use was the default, the *de re* reading of a belief report comes out as a default, because it corresponds to the strongest intentions and the strongest intentionality. So, Figure 1 comprises three possible representations (three possible compositionality-intentionality mergers).

In addition to definite descriptions in extensional and in propositional attitude contexts, the mechanism of the principles of Default Semantics has been applied to a variety of language expressions and constructions, including proper names (Jaszczolt, 1999b), presuppositional expressions (Jaszczolt, 2002a, 2002b), expressions of temporality and modality, and tentatively to numerals and sentential connectives (Jaszczolt, in press b, in press c). Naturally, the PI principle will not always be relevant. The referential intention will not always be present, and even when it is, it may not pertain to the assessment of the default or nondefault status of various readings. For example, in an assessment of the default meaning of *will* from among the epistemic necessity *will* in (3), dispositional necessity *will* in (4), and a marker of future tense in (5), it is the intention to inform the addressee about a certain content that is graded from the strongest to the weakest:

- (3) Mary will be in the opera now.
- (4) Mary will sometimes go to the opera in her tracksuit.
- (5) Mary will go to the opera tomorrow night.

The Default-Semantic account of *will* also demonstrates that modal and temporal senses of *will* are traceable to one, overarching modal concept (akin to the sentential operator of *acceptability* in Grice, 2001). And since *will* is modal, it follows that the assignment of defaults has to be reversed as compared with the examples previously discussed: the weakest intentionality corresponds to the default sense of *will*, and this, predictably, turns out to be the regular future marker in (5) (for a formal account, see Jaszczolt, in press a).

Not all default interpretations are reducible to cognitive defaults. For example, the interpretation of possessives, as in (6), is dependent on the addressee's background knowledge and the context, rather than on the properties of mental states.

- (6) Peter's book is about a glass church.

Similarly, inferences to a stereotype ('female nurse'), such as in (7), are not the case of the strength of intending but rather stem from the acquaintance with social and cultural practices.

- (7) They employed a nurse to look after the patient.

Such default interpretations belong to the category of social and cultural defaults and are not always of central interest to semantic theory.

The phenomenon of negative-raising, i.e., the tendency for negation on the main clause to be interpreted as negation on the subordinate clause, is not an obvious cognitive default, but here we must be cautious. Neg-raising unpredictably applies to some relevant verbs but not to others, as (8) and (9) demonstrate.

- (8) I don't think he is dishonest. (communicates, defeasibly: 'I think he is not dishonest.')
- (9) I don't hope he will win. (does not communicate: 'I hope he will not win.')

The important question at this point is to ask about the scope of applicability of the theory. The question of the scope of applicability can be taken in the narrow and in the wide sense. In the narrow sense, we ask which default interpretations can be regarded as cognitive defaults, traceable to the properties of mental states. Cognitive defaults are rather widespread. In addition to the examples already mentioned, numerals seem to default to the 'exactly' meaning, rather than being underdeterminate between 'at least,' 'at most,' and 'exactly,' or having an 'at least' semantics. The enrichment of some sentential connectives such as *if* (to 'if and only if') and *or* (to exclusive *or*) can possibly also be traced to the strength of the informative intention and intentionality. This proposal concerning connectives and numerals is still highly programmatic and in need of further research. It is signaled here in order to shed some light on possible applications of cognitive defaults. In the wide sense, Default Semantics also comprises social and cultural defaults simply by assigning them an epistemological status that has nothing to do with the compositionality-intentionality merger.

To sum up: Default Semantics postulates a level of utterance interpretation called a compositionality-intentionality merger and thereby significantly decreases the role of underspecification in semantic theory. It distinguishes cognitive defaults and intention-based degrees of departures from these defaults, triggered by the addressee's knowledge base and the context. The theory also acknowledges the existence of social and cultural defaults whose source lies beyond semantics proper.

See also: Compositionality: Semantic Aspects; *De Dicto* versus *De Re*; Descriptions, Definite and Indefinite; Philosophical Aspects; Discourse Representation Theory; Implicature; Presupposition; Referential versus Attributive; Semantics–Pragmatics Boundary.

Bibliography

- Asher N & Lascarides A (2003). *Logics of conversation*. Cambridge: Cambridge University Press.
- Bach K (1994). 'Semantic slack: what is said and more.' In Tsohatzidis S L (ed.) *Foundations of speech act theory: philosophical and linguistic perspectives*. London: Routledge. 267–291.
- Carston R (1988). 'Implicature, explicature, and truth-theoretic semantics.' In Kempson R M (ed.) *Mental representations: the interface between language and reality*. Cambridge: Cambridge University Press. 155–181.
- Carston R (2002). *Thoughts and utterances: the pragmatics of explicit communication*. Oxford: Blackwell.
- Grice H P (1975). 'Logic and conversation.' In Cole P & Morgan J L (eds.) *Syntax and semantics, vol. 3*. New York: Academic Press. Reprinted in Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press. 22–40.
- Grice P (2001). Warner R (ed.) *Aspects of reason*. Oxford: Clarendon Press.
- Husserl E (1900–1901). *Logical investigations (vol 2)*. Findlay J N (trans.) (1970; reprinted in 2001). London: Routledge and Kegan Paul.
- Jaszczolt K M (1999a). 'Default semantics, pragmatics, and intentions.' In Turner K (ed.) *The semantics/pragmatics interface from different points of view*. Oxford: Elsevier Science. 199–232.
- Jaszczolt K M (1999b). *Discourse, beliefs, and intentions: semantic defaults and propositional attitude ascription*. Oxford: Elsevier Science.
- Jaszczolt K M (2000). 'The default-based context-dependence of belief reports.' In Jaszczolt K M (ed.) *The pragmatics of propositional attitude reports*. Oxford: Elsevier Science. 169–185.
- Jaszczolt K M (2002a). 'Against ambiguity and underspecification: Evidence from presupposition as anaphora.' *Journal of Pragmatics* 34, 829–849.
- Jaszczolt K M (2002b). *Semantics and pragmatics: meaning in language and discourse*. London: Longman.
- Jaszczolt K M (in press a). 'Futurity in Default Semantics.' In von Heusinger K & Turner K (eds.) *Where semantics meets pragmatics*. Oxford: Elsevier.
- Jaszczolt K M (in press b). 'Prolegomena to Default Semantics.' In Marmaridou S, Nikiforidou K & Antonopoulou E (eds.) *Reviewing linguistic thought: converging trends for the 21st century*. Berlin: Mouton.
- Jaszczolt K M (in press c). *Default semantics: Foundations of a compositional theory of acts of communication*. Oxford: Oxford University Press.
- Kamp H & Reyle U (1993). *From discourse to logic: introduction to model-theoretic semantics of natural language, formal logic and Discourse Representation Theory*. Dordrecht: Kluwer.
- Levinson S C (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Reyle U (1993). 'Dealing with ambiguities by underspecification: construction, representation and deduction.' *Journal of Semantics* 10, 123–179.

Definite and Indefinite

B Abbott, Michigan State University, East Lansing, MI, USA

© 2006 Elsevier Ltd. All rights reserved.

What Does 'Definite' Mean?

'Definite' and 'indefinite' are terms that are usually applied to noun phrases (NPs). In English, *the* is referred to as 'the definite article,' and *a/an* as 'the indefinite article.' Noun phrases (NPs) that begin with *the* (e.g., *the Queen of England, the book*), which are also called (especially in the philosophical literature) 'definite descriptions,' are generally taken to be prototypical examples of definite NPs in English. However, it should be noted that not all of them show the same pieces of behavior that have come to be taken as criterial for definiteness. Similarly, NPs that begin with *a/an*

(*an elephant, a big lie*), 'indefinite descriptions,' are prototypical examples of indefinite NPs. (Plural indefinite descriptions use the determiner *some*.)

Uniqueness?

Exactly what differentiates definite from indefinite NPs has been a matter of some dispute. One tradition comes from the philosophical literature, more specifically Bertrand Russell's classic work on denoting phrases (Russell, 1905). On this tradition, what distinguishes *the* from *a/an* is uniqueness – more specifically the existence of one and only one entity meeting the descriptive content of the NP. So, while use of an indefinite description in a simple, positive sentence merely asserts existence of an entity meeting the description, use of a definite description asserts in addition its uniqueness in that regard. While (1a), on this

view, is paraphrasable as (1b), (2a) is equivalent to (2b).

- (1a) I met an owner of El Azteco.
- (1b) There is at least one owner of El Azteco whom I met.
- (2a) I met the owner of El Azteco.
- (2b) There is one and only one owner of El Azteco, and I met that individual.

It should be noted that Russell was concerned to capture the meaning of definite descriptions in a formal language of logic. Also, on his analysis both definite and indefinite descriptions are quantificational expressions (like explicitly quantified NPs, such as *every apple*, *no unwanted visitors*). The idea that definite descriptions are quantificational has been questioned by others, who view these NPs instead as referential. Fewer people question the idea that indefinite descriptions are quantificational, although some (primarily linguists, rather than philosophers) assume that they, too, are referential.

The uniqueness theory seems to accord well with our intuitions. It also is supported by the fact that when we stress the definite article contrastively, it brings out the sense of uniqueness. Example (3) seems to be inquiring as to whether there is more than one owner, or only one.

- (3) Did you meet an owner of El Azteco or the owner?

It might seem that this approach would necessarily be confined to singular NPs. However, as argued by Hawkins (1978), the notion of uniqueness can be extended to plurals by employing the idea of exhaustiveness – the denotation of a definite consists of everything meeting the descriptive content of the NP. An NP like *the owners of El Azteco* would thus be held to be very similar to *all the owners of El Azteco*.

The first challenge to Russell's analysis of definite descriptions was put forward by P. F. Strawson, who argued that sentences containing definite descriptions are not used to assert the existence and uniqueness of an entity meeting the descriptive content in question. Instead, Strawson argued, definite descriptions are referential NPs, and the existence and uniqueness of a referent is presupposed (cf. Strawson, 1950; in this seminal work, Strawson did not use the term 'presuppose,' although it appeared very quickly in reference to the phenomenon in question). Strawson also argued that if the presupposition fails, the sentence as a whole is neither true nor false. Thus, in the case of (2a), should it turn out that no one owns El Azteco (perhaps it is a government installation), an addressee of (2a) would not respond "That's false!" but would correct the speaker's mistaken presupposition.

Another, more serious problem for Russell's analysis has attracted a lot of attention more recently, and that is the fact that in a great number of cases, perhaps the vast majority, the descriptive content of a definite description is not sufficient to pick out a unique referent from the world at large. One example of such an 'incomplete description' is in (4).

- (4) Please put this on the table.

The sentence in (4) is readily understandable despite the fact that the world contains millions of tables. There are two main kinds of approach to dealing with this problem. A syntactic solution would propose that there is sufficient additional descriptive material tacitly present in the NP – e.g., *the table next to the armchair in the living room of the house at 76 Maple Avenue, Eastwood, Kansas, USA*. But it would be hard to explain how an addressee would guess which descriptive content had been left tacit. On a more plausible approach, the uniqueness encoded in definite descriptions should be understood relative to a context of utterance, which would include only those items in the surroundings of the discourse participants and those items mentioned in the course of the conversation or understood to be relevant to its topic. However, this runs into a problem with examples like (5), first pointed out by James McCawley (McCawley, 1979).

- (5) The dog got into a fight with another dog.

David Lewis proposed that definite descriptions denote the most salient entity meeting the descriptive content (Lewis, 1979).

Familiarity?

The other main tradition concerning the meaning of definiteness generally cites the Danish grammarian Paul Christophersen. In Christophersen's view, what distinguishes definite from indefinite descriptions is whether or not the addressee of the utterance is presumed to be acquainted with the referent of the NP. In an often cited passage, Christophersen remarked: "Now the speaker must always be supposed to know which individual he is thinking of; the interesting thing is that the *the*-form supposes that the hearer knows it too" (Christophersen, 1939: 28). This approach appears to fare better with examples like (4), where, indeed, it seems that the speaker must be assuming that the addressee knows which table the speaker is referring to.

Within current linguistic theory, the familiarity approach was revived by the work of Irene Heim (1982, 1983). Like Strawson, Heim argued that definite descriptions are referential rather than quantificational; however, she also argued that indefinite

descriptions are referential as well. Heim took the uses of definite and indefinite descriptions as they occur in (6) as typifying their semantics.

- (6) Mary saw a movie last week. The movie was not very interesting.

In the mini discourse in (6), the indefinite NP *a movie* is used to introduce a new entity into the discourse context. Subsequently that entity is referred to with a definite (*the movie*). Notice that we might as easily have referred to the movie in the second sentence of (6) with a pronoun: ...*It was not very interesting*. Heim grouped pronouns and definite descriptions together as being governed by a “Familiarity” condition: use of a definite is permitted only when the existence of the referred-to entity has been established in the particular discourse. Indefinite descriptions, on the other hand, are subject to a “Novelty” condition: they presuppose that their referent is being introduced into the discourse for the first time. It’s easy to see that this will solve the problem of incomplete descriptions. An example like (5) would be used only when the first dog referred to was presumed to be known to the addressee.

Though the familiarity theory is very plausible for a number of uses of definite descriptions, there are some kinds of cases it does not appear to cover very well. One of these is definite descriptions where the descriptive content of the NP is sufficient to determine a unique referent, no matter what the context. Some examples are given in (7).

- (7a) Mary asked the oldest student in the class to explain everything.
(7b) Philip rejected the idea that languages are strictly finite.

Here we need not assume that the addressee is familiar with the referents of the underlined NPs or that these referents had been mentioned previously in the conversation. Note, too, that in this kind of case, the indefinite article is not allowed, as shown in (8). (The asterisk in front of these examples indicates that they are not well formed.)

- (8a) * Mary asked an oldest student in the class to explain everything.
(8b) * Philip rejected an idea that languages are strictly finite.

And even when the descriptive content is not sufficient to determine a unique referent relative to the whole world, there are examples where the content may determine a unique referent in context. In these cases, too, the definite article may be used, even if the addressee is not assumed to know who or what is being talked about. An example is given in (9).

- (9) Sue is mad because the realtor who sold her house overcharged his fee.

Adherents to the familiarity theory often invoke the idea of accommodation (following Lewis, 1979) to explain these uses. The idea is that addressees are willing to accept a definite description if they are able to figure out the intended referent.

Some Puzzling Cases

While most occurrences of definite descriptions are consistent with both the uniqueness theory and the familiarity theory, there are several kinds that don’t match either theory. One group of examples is given in (10).

- (10a) Horace took the bus to Phoenix.
(10b) The elevator will take you to the top floor.

It seems that with modes of transportation, a singular definite description can be used despite the fact that there are, e.g., many buses to Phoenix, and the building in (10b) may have many elevators. We also don’t suppose that the addressee will be familiar with the particular bus or elevator in question. A different kind of case is illustrated in (11).

- (11a) My uncle wrote something on the wall.
(11b) We camped by the side of a river.
(11c) She shot herself in the foot.

These sentences are well formed even though rooms typically have more than one wall, rivers more than one side, and people more than one foot. It may be relevant that these are locations. In all of these cases, as pointed out by Du Bois (1980), to use an indefinite article puts too much emphasis on the location, as though it were inappropriately being brought into focus.

A third kind of example shows some dialectal variation. The example in (12) is well formed in American English, although in British English the article would be missing from the underlined NP.

- (12) My mother is in the hospital.

Compare the examples in (13), which are good in both dialects.

- (13a) Bill went to school this morning.
(13b) If you’re very good, you’ll get to heaven some day.

The examples in (14), also good for English in general, indicate a certain amount of idiomaticity.

- (14a) I heard it on the radio.
(14b) I saw it on TV.

It seems that some nouns simply require the definite article, while others are fine without it.

Finally, some adjectives call for the definite article in English, despite their not restricting the reference of NPs they occur in to either a unique or a familiar referent.

- (15) She gave the wrong answer and had to be disqualified.

It is not clear whether these examples indicate the need for a brand-new theory of the definite article in English or are just idiomatic exceptions to the rule.

Grammatical Phenomena

Sensitivity to definiteness of NP is called a definiteness effect, and a number of constructions are believed to have such an effect.

Existential Sentences

One of the earliest constructions showing a definiteness effect to be noticed was existential, or *there be*, sentences. Examples like those in (16) are quite natural, but the corresponding sentences in (17) sound peculiar, if not downright ungrammatical.

- (16a) There is a book in the shop window.
 (16b) There were some bachelors on board the ship.
 (17a) * There is the book in the shop window.
 (17b) * There were the bachelors on board the ship.

One complicating factor is the existence of a construction that is similar to the existential construction but that is used in a restricted set of circumstances. The latter kind of sentence, often called a list existential, typically seems to be used to offer entities to fulfill some role or purpose. However, this kind of existential does not allow a locative prepositional phrase to follow the focus NP. Examples just like those in (17) but where the prepositional phrase is an NP modifier, as in (17'), could be used in reply to the questions in (18).

- (17'a) There is the book in the shop window.
 (17'b) There were the bachelors on board the ship.
 (18a) What can we get for Bill for his birthday?
 (18b) Weren't there any people around to help?

The more common type of existential, like those in (16), can be called locative existentials. In this case, the prepositional phrase that follows the focus NP is a separate constituent that locates the item in question. It is only locative conditionals that show a definiteness effect, and these have been used as a test for definiteness, as we see in the section 'Other Kinds of Definite and Indefinite NPs.'

The *Have* Construction

Another construction, similar to existential sentences, is one involving the verb *have* when it is used to indicate inalienable possession. Here, too, we see a definiteness effect, in that the examples in (19) are natural, while those in (20) are not.

- (19a) She had a full head of hair.
 (19b) He had a sister and two brothers.
 (20a) * She had the full head of hair.
 (20b) * He had the sister and the two brothers.

It is perhaps not too surprising that these two constructions should show a similar definiteness effect, since *have* and *be* verbs are often used for similar kinds of propositions in the world's languages.

Other Kinds of Definite and Indefinite NPs

As shown in the first section, there is no commonly agreed on essence of definiteness or indefiniteness. Hence the need for some kind of diagnostic for these properties. Ability to occur naturally in a locative existential has become the main diagnostic used.

Other Kinds of Definite NPs

As noted in this article, Heim assumed that pronouns are definite, like definite descriptions. Others agree with this categorization. And as we might expect, pronouns do not go naturally in locative existentials. The sentences in (21) are not natural.

- (21a) * There was it in the fireplace.
 (21b) * There were them all over the floor.

Pronouns seem to fit both the uniqueness and the familiarity conceptions of definiteness. When they are used, it is assumed that there is a unique intended referent within the discourse context, and it is also assumed that the addressee will know who or what the speaker was intending to refer to.

Another subcategory of NP that is typically assumed to be definite consists of proper names. Like pronouns and definite descriptions, these do not occur naturally in locative existentials.

- (22a) * There was Joan in the library.
 (22b) * There is France in the United Nations.

Although it might not be so obvious as it is with pronouns, proper names also seem definite by both theories of definiteness. Proper names behave as though they have a unique referent; they cannot accept restrictive adjectives or other restrictive modifiers. And in fact in most contexts, each proper name does have a unique referent. On the other hand, we do not usually use a proper name unless we assume that our addressee has already been introduced to the referent.

A third kind of NP that is generally agreed to be definite would be those that have a demonstrative determiner: *this*, *that*, *these*, or *those*. These cannot occur naturally in a locative existential, as shown (23) and (24).

- (23a) * There was that book over there in Mary's bag last Tuesday.
 (23b) * There are these applicants waiting to see the dean.

In addition, NPs with a possessive determiner are usually classed as definite.

- (24) * There was Mary's car in the driveway.

Indeed, NPs with possessive determiners are typically regarded as belonging to the category of definite descriptions.

Some kinds of quantified NPs cannot occur naturally in existential sentences, and this has led some people to consider them to be definite NPs. Some examples are given in (25).

- (25a) * There were all the students at the party.
 (25b) * There were most red buttons on the dress.

However, it is possible that these NPs should not be classified as definite and that there is some other reason that they cannot occur in locative existential sentences.

Bare NPs

One interesting kind of NP in English has received a significant amount of attention. So-called bare NPs do not have any determiner, and the head noun must be either plural or a mass noun. These NPs have (at least) two distinct uses. Sometimes they are interpreted generically, as in (26).

- (26a) Mary likes sharpened pencils.
 (26b) Water with fluoride in it is good for the teeth.

The sentences in (26) concern the whole category referred to by the underlined NP. On the other hand, sometimes these bare NPs have an existential interpretation, in that they are referring to just some members or a subpart of the category.

- (27a) Mary bought sharpened pencils.
 (27b) There was water with fluoride in it in the test tube.

As can be seen in (27b), when bare NPs occur in a locative existential sentence, they can have only the existential interpretation, and not the generic one.

Other Types of Indefinite NPs

In addition to indefinite descriptions, and bare NPs with the existential interpretation, there are other

types of NPs that go naturally in locative existentials. Some examples of quantified NPs are shown in (28).

- (28a) There are a few pieces of cake left.
 (28b) There were few, if any, freshpersons at the school fair.
 (28c) There are many problems for that course of action.
 (28d) There are some big flecks of paint on the back of your coat.

If we use natural occurrence in a locative existential as a diagnostic, then these other types would also be classified as indefinite NPs.

In addition to these, there are some other unexpected cases of NPs that look as though they should be definite, because they have definite determiners, but that can appear naturally in a locative existential. One kind, noticed first by Prince (1981), uses the proximal demonstrative determiner (*this*, *these*), but with an indefinite reference, as in (29).

- (29a) There was this strange note on the blackboard.
 (29b) There are these disgusting globs of stuff in the bowl.

Note that this is definitely a different use of these determiners. The examples in (29) would not be used with any kind of pointing gesture, and indeed, they could be used in a phone conversation, where the addressee is not in the same perceptual situation as the speaker. Also, it is worth noting that this indefinite use of *this* and *these* is somewhat marked stylistically. Examples like those in (29) would not appear in a formal context.

Finally, there are some kinds of NPs that look like definite descriptions, but whose sense is indefinite, and that can appear naturally in existentials.

- (30a) There was the nicest young man at the picnic!
 (30b) There were the same nominees on both ballots.

As Prince (1992) pointed out, an NP like *the same nominees* can have two interpretations. One is anaphoric, as in (31).

- (31) The Executive Committee came up with a list of nominees, and it happened that the Nominating Committee chose the same nominees.

Here *the same nominees* refers back to the Executive Committee's list and means that the Nominating Committee's choices were the same as the Executive Committee's. On this interpretation, *the same nominees* would be definite. However, the interpretation in (30b) is different: it means that the two ballots had the same choices. This interpretation is apparently indefinite, given the ability of this NP to occur naturally in the existential sentence in (30b).

Other Kinds of Categorizations

A simple binary distinction like definite vs. indefinite may be too crude, especially if we are trying to classify NPs in general. Furthermore, it may be more useful to look at the role of NP form with respect to discourse function. A number of researchers have turned to the idea of information status – an extension of the familiarity idea, but with greater articulation.

Old and New

Prince (1992) argued that we need to distinguish two ways in which information can be novel or familiar, new or old. One is with respect to (the speaker's assumption about) the addressee, which Prince called Hearer-old and Hearer-new. The speaker assumes that the addressee is already acquainted with the referent of a Hearer-old NP, whereas Hearer-new NPs are assumed to introduce new entities to the addressee. On the other hand, entities can be new or old with respect to a discourse: Discourse-old or Discourse-new. Discourse-old NPs refer to entities that have already been mentioned in the current discourse, in contrast to Discourse-new NPs. Prince found that it was the category of Hearer-old/Hearer-new that correlated roughly with the definite/indefinite distinction, rather than Discourse-old/Discourse-new. This seems to agree more with Christophersen's than with Heim's conception of definiteness and indefiniteness.

The Givenness Hierarchy

Gundel *et al.* (1993) proposed a hierarchy of givenness corresponding to the degree to which the referent of an NP is assumed to be cognitively salient to the addressee. Each point in the hierarchy corresponds to one or more NP forms. At one end of the hierarchy is the weakest degree of knownness, which Gundel *et al.* labeled 'type identifiable.' This end corresponds to indefinite descriptions, and the criterion for their use is just that the addressee be familiar with the kind of thing denoted. At the other extreme, we find NPs denoting entities that are currently 'in focus,' and pronouns require this extreme of cognitive salience.

Definite descriptions are about midway in the hierarchy, requiring unique identifiability for their referents. Just to the weaker side are the indefinite *this/these* NPs. On the more salient side of definite descriptions are NPs with demonstratives (*this, that, these, those*) as determiners.

One special aspect of this treatment is that the criteria for each place on the hierarchy are increasingly stringent and subsume the criteria for all less stringent points; that is, the hierarchy is an implicational

one. Hence, indefinite descriptions, which have the weakest requirement, can appear anywhere in principle – even when their referents are in focus. However, general conversational principles militate against using an indefinite description in such a situation, as it would be misleading in suggesting that only the weakest criterion had been satisfied.

The Accessibility Hierarchy

A third approach, similar to the one just mentioned but with its own distinct characteristics, was developed by Mira Ariel (1990, 2001). Ariel proposed an even more articulated accessibility hierarchy, reflecting the marking of NPs according to how accessible in human memory their referents should be. Upwards of 15 distinct categories are represented, ranging from full name plus modifier (at the least accessible end) to zero pronouns (represented with \emptyset), which are found in constructions like those in (32).

- (32a) Mary wanted \emptyset to build a cabin.
 (32b) Open \emptyset in case of fire.

Full name plus modifier (e.g., *Senator Hillary Clinton*) is distinguished from full name, last name alone, or first name alone, each of which receives a separate spot on the hierarchy. Similarly long definite descriptions, with a lot of descriptive content, are distinguished from short ones, and stressed pronouns from unstressed pronouns. (The list does not contain quantified and [other] indefinite NPs, which, as noted in this article, are often considered not to be referential expressions.)

Ariel's claim was that the hierarchy of NP forms corresponds to accessibility, where the latter is determined by a number of factors, including topic-hood, recency and number of previous mentions, and how stereotypic the referent is for the context. The NP forms go generally from fullest and most informative to briefest and least informative, \emptyset being the limiting case. The idea is that an NP form typically encodes an appropriate amount of information for the addressee to achieve easy identification of the referent.

Definite and Indefinite in Other Languages

The examples used so far in this article have been taken from English. However, many other languages have definite and/or indefinite articles, though by no means all of them. Lyons (1999) described the explicit marking of definiteness – whether with an article or a nominal inflection – as an areal feature that characterizes the languages of Europe and

the Middle East in particular, although it can be found elsewhere in the world as well (Lyons, 1999: 48). Definite articles often seem to develop out of demonstrative determiners, as was the case in English. Indefinite articles, on the other hand, often come from the word for 'one.'

Some languages, e.g., Irish (Irish Gaelic), have only a definite article, whereas others, e.g., Turkish, mark only indefinites explicitly. The examples in (33) and (34) are taken from Lyons (1999: 52).

- (33) *Irish*
 (33a) an bord 'the table'
 (33b) bord 'a table'
- (34) *Turkish*
 (34a) ev 'house', 'the house'
 (34b) bir ev 'a house'

Even among languages that have both definite and indefinite marking, the usages typically do not match exactly across languages. Thus, bare NPs in English have a generic use (as in the examples in [26]). French also has both definite and indefinite determiners but, unlike English, would use the definite determiner in examples like those in (26).

- (35) *French*
 (35a) Marie aime les crayons bien taillés.
 (35b) L'eau au fluor est bonne pour les dents.

In languages that do not use articles or some other explicit marking for definiteness or indefiniteness, word order may affect interpretation in that way, as in the examples in (36) and (37), from Chinese and Russian.

- (36) *Mandarin Chinese*
 (36a) Zhuo-zi shang you shu.
 table on have book
 'There is a book (or books) on the table.'
 (36b) Shu zai zhuo-zi shang.
 book is located table on
 'The book is on the table.'
- (37) *Russian*
 (37a) Na stolé lezhít karta.
 on table lies map
 'There is a map lying on the table.'
 (37b) Karta lezhít na stolé.
 map lies on table
 'The map is lying on the table.'

However, it should be noted that word order variation also interacts with topicality and the distribution of new and old information in a sentence and that this affects the definiteness or indefiniteness of an NP's interpretation.

For a full discussion of the expression of definiteness and indefiniteness in a variety of the world's languages, see Lyons (1999).

See also: Descriptions, Definite and Indefinite: Philosophical Aspects; Presupposition; Proper Names: Philosophical Aspects.

Bibliography

- Abbott B (2004). 'Definiteness and indefiniteness.' In Horn L R & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 122–149.
- Ariel M (1990). *Accessing noun phrase antecedents*. London: Routledge.
- Ariel M (2001). 'Accessibility theory: an overview.' In Sanders T, Schliperoord J & Spooren W (eds.) *Text representation*. Amsterdam and Philadelphia: John Benjamins. 29–87.
- Birner B J & Ward G (1998). *Information status and noncanonical word order in English*. Amsterdam and Philadelphia: John Benjamins.
- Carlson G N (1977). 'A unified analysis of the English bare plural.' *Linguistics and Philosophy* 1, 413–456.
- Christophersen P (1939). *The articles: a study of their theory and use in English*. Copenhagen: Munksgaard.
- Diesing M (1992). *Indefinites*. Cambridge, MA: MIT Press.
- Du Bois J W (1980). 'Beyond definiteness: the trace of identity in discourse.' In Chafe W L (ed.) *The pear stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex. 203–274.
- Gundel J K, Hedberg N & Zacharski R (1993). 'Cognitive status and the form of referring expressions in discourse.' *Language* 69, 274–307.
- Haspelmath M (1997). *Indefinite pronouns*. Oxford: Oxford University Press.
- Hawkins J A (1978). *Definiteness and indefiniteness*. Atlantic Highlands, NJ: Humanities Press.
- Hawkins J A (1991). 'On (in)definite articles: implicatures and (un)grammaticality prediction.' *Journal of Linguistics* 27, 405–442.
- Heim I (1982). The semantics of definite and indefinite noun phrases. Ph.D. diss., University of Massachusetts.
- Heim I (1983). 'File change semantics and the familiarity theory of definiteness.' In Bauerle R, Schwarze C & von Stechow A (eds.) *Meaning, use and the interpretation of language*. Berlin: Walter de Gruyter. 164–189.
- Lewis D (1979). 'Scorekeeping in a language game.' *Journal of Philosophical Logic* 8, 339–359.
- Löbner S (1985). 'Definites.' *Journal of Semantics* 4, 279–326.
- Lumsden M (1988). *Existential sentences: their structure and meaning*. London: Croom Helm.
- Lyons C (1999). *Definiteness*. Cambridge: Cambridge University Press.
- McCawley J D (1979). 'Presupposition and discourse structure.' In Oh C-K & Dinneen D (eds.) *Syntax and semantics, vol. 11: Presupposition*. New York: Academic Press. 371–388.
- Milsark G (1977). 'Toward an explanation of certain peculiarities of the existential construction in English.' *Linguistic Analysis* 3, 1–29.
- Neale S (1990). *Descriptions*. Cambridge, MA: MIT Press.

- Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: MIT Press.
- Prince E F (1981). 'On the inferencing of indefinite *this* NPs.' In Joshi A K, Webber B L & Sag I A (eds.) *Elements of discourse understanding*. Cambridge: Cambridge University Press. 231–250.
- Prince E F (1992). 'The ZPG letter: subjects, definiteness, and information status.' In Mann W C & Thompson S A (eds.) *Discourse description: diverse linguistic analyses of a fund-raising text*. Amsterdam and Philadelphia: John Benjamins. 295–326.
- Reimer M & Bezuidenhout A (eds.) (2004). *Descriptions and beyond*. Oxford: Oxford University Press.
- Reuland E J & ter Meulen A G B (eds.) (1987). *The representation of (in)definiteness*. Cambridge, MA: MIT Press.
- Roberts C (2003). 'Uniqueness in definite noun phrases.' *Linguistics and Philosophy* 26, 287–350.
- Russell B (1905). 'On denoting.' *Mind* 14, 479–493.
- Strawson P F (1950). 'On referring.' *Mind* 59, 320–344.
- Woisetschlaeger E (1983). 'On the question of definiteness in "an old man's book".' *Linguistic Inquiry* 14, 137–154.

Definitions: Uses and Varieties of

G Longworth, Birkbeck College, University of London, London, England, UK

© 2006 Elsevier Ltd. All rights reserved.

Uses

'Definition' is the activity of explaining to an audience the meaning of an expression. 'A definition' is a product of that activity: a sentence the understanding of parts of which (the part of the sentence providing explanation, the *definiens*) can underwrite an audience's understanding of the rest (the part of the sentence being defined, the *definiendum*). For example, understanding 'is the activity of explaining the meaning of an expression' (*definiens*) might enable one to understand a meaning of 'definition' (*definiendum*). Notice that 'definition' needn't proceed via 'definitions.' Perhaps the first explanations of meaning to which a child is exposed don't come via sentences.

Besides the immediate purpose of underwriting explanations of meaning, a definition can serve countless others. One may stipulate that an expression is to enjoy a meaning – deploying a 'stipulative definition.' E.g., for purposes of this entry, let *x* be a *definition* if *x* is a sentence used to explain meaning. (Here and throughout, initial universal quantification is suppressed and use/mention distinctions ignored.) Other purposes of stipulation include abbreviation – hence, 'abbreviative definition' – itself at the service of tractability and comprehensibility – and marking out the *definiens* as of special interest, perhaps as carving at an important joint (Whitehead and Russell, 1910).

An alternative purpose is to describe the meaning an expression carries with respect to some language or population – a 'descriptive' or 'lexical definition.' Thus, in English, *x* is a *definition* iff *x* is a sentence used to explain meaning. Less immediate purposes here include illuminating a less well understood *definiendum* by appeal to a better understood *definiens*;

revealing the basis of one's understanding of the *definiendum*; or establishing dependence of the *definiendum* on the *definiens*. But the basic purpose of descriptive definition – explaining the meaning of the *definiendum* – is independent of the viability of these other purposes. This is good, since it would be surprising if many expressions in use were redundant.

A third purpose is 'explication' or 'explicative definition.' Here one stipulates with the aim of approximating to the meaning of an ordinary expression. The aim is to balance two requirements: first, the new expression should be fit to do duty for the old, at least for some purposes; and second, the new expression should improve upon the ordinary along some valued dimension, perhaps clarity or consistency (Carnap, 1928, 1947). Explication is risky, as it is in general impossible to specify in advance the range of important duties an expression must perform. The definitions recently presented are sufficiently vague and ambiguous to meet the first requirement, if not the second.

Whatever one's purposes, the capacity of a definition to serve them is relative to the context (circumstances) in which the definition is offered. In particular, it is relative to the needs of one's audience and to their capacities and informational situation. The role of audience capacities and collateral information is difficult to articulate in detail, but can be illustrated. Someone who lacks the capacities needed to understand 'explain' will not gain understanding of 'definition' from the definition offered above. Moreover, it's plausible that, since true synonymy (sameness of meaning) is rare, most dictionary definitions rely heavily on audiences' knowledge and abilities, often supplying little more than hints from which the intellectually privileged are able to derive understanding.

Mention of contextual features is often suppressed, especially in logic. Suppression is motivated by aims, such as balancing maximal generality against formal tractability. Aiming for generality induces logicians to

articulate assumptions and rely only on capacities widely possessed amongst thinkers. Seeking tractability induces restrictions on the range of uses of an expression that a definition is required to explain. Logicians typically require definitions to convey meaning to audiences competent with the logical apparatus and language of their logical theory, without relying on special capacities or features of the circumstances in which the definition is offered. It doesn't follow that definitions offered by logicians are more than comparatively context-free. Neither does it follow that explanations of meaning outside of logic are required to attain a similar level of context-freedom.

Varieties

Having seen something of the assortment of uses of definitions, we can consider some varieties of definitions. There are as many forms of definition as there are ways of using sentences to enable someone to discover the meaning of an expression. Those listed below are included for their exhibition of variety, and also because they illustrate the context sensitivity of definition. This is addressed more explicitly in the section 'Uses Again.'

Comparatively Context-Free Forms of Definition

Explicit definition involves assuming an audience to understand the *definiens* in advance, and presenting the *definiendum* as something that can replace the *definiens* for current purposes. For example,

- (1) A **brother** is a male sibling

Here, an audience is informed that *brother* can be used wherever *male sibling* is used. An explicit definition accomplishes this by associating with the *definiendum* an expression that can serve in function-preserving replacements for that expression, perhaps through synonymy, or some weaker equivalence.

An interesting form of quasi-explicit definition is 'recursive' or 'inductive definition.' (Quasi-explicit since it fails to provide an independent expression that can be used wherever the expression to be defined can be used.) A recursive process is one that takes its own outputs as inputs, generating new outputs that can serve as inputs, and so forth. Use of recursive definitions enables us to characterize the meaning – e.g., extension – of expressions when that meaning can, or can only, be generated by a recursive process. For example:

- (2) x is a **direct ancestor** of y iff x is a parent of y or x is a parent of a direct ancestor of y

Here, the *definiendum* appears in the *definiens* so that the extension of the *definiens* cannot be determined in advance of partial determination of the extension of

the *definiendum*. This in turn cannot be determined in advance of partial determination of the extension of the *definiens*. This is apt to seem viciously circular, but it isn't. Vicious circularity is avoided because the basis clause, x is a parent of y , affords a means to start the recursive process independently of grasp of the meaning of x is a direct ancestor of y . The 'inductive step' – or x is a parent of a direct ancestor of y – can then make use of the output of the basis clause to generate a new input, and so forth.

Explicit definition is unhelpful when the framework in which it is given deploys the expression to be defined, or when an audience lacks other expressions able to do the same work. In such cases, one might deploy 'implicit (or contextual) definition.' An implicit definition explains an expression's meaning through appeal to other elements in the definition. But unlike an explicit definition, the other elements in an implicit definition need not be equivalent to the *definiendum*. Implicit definition involves stipulating the truth of sentences involving the *definiendum* in a way that fixes its meaning as the unique meaning able to sustain the truth of the sentences so stipulated.

One example was Bertrand Russell's account of definite descriptions, sentences of the form *The F is G* like *The King of France is bald*. Rather than presenting an explicit definition of *The* or *The F*, Russell explicitly defined whole sentences in which they occur, and thereby implicitly defined them, via (3):

- (3) $\text{The } F \text{ is } G \text{ iff } (\exists x) (Fx \ \& \ (\forall y) (Fy \supset x = y) \ \& \ Gy)$

The right-hand side reads: there is exactly one F and every F is G (Russell, 1905).

Another example involves the use of a definite description to explain a proper name's reference. The reference of *Jack the Ripper* might be explained using the following sentence:

- (4) **Jack the Ripper** is the perpetrator of the Whitechapel murders

Although (4) can be used to explain the meaning of *Jack the Ripper*, it does so without identifying it with the meaning of the descriptive phrase *the perpetrator of the Whitechapel murders*. Proper names – unlike descriptive phrases – are 'rigid designators.' They refer, crudely, to the same object in every possible world. So, while *the perpetrator of the Whitechapel murders* denotes different individuals in different possible worlds – depending on who in those worlds committed the crimes – *Jack the Ripper* refers to the same person in each world: whoever committed the crimes in the actual world. It follows that (4) is only 'contingently' true (might have been false). But, arguably, since (4) is stipulated, it is knowable *a priori* (Kripke, 1980; Evans, 1979; Soames, 2003:

397–422). This effect is mediated by the audience's standing competence with the category of proper names. So this feature of context plays a role in mediating the transition from what is presented in a definition to the understanding conveyed.

Comparatively Context-Dependent Definitions
More obviously context-dependent forms of definition involve appeal to examples, and so to the classificatory abilities of one's audience.

Ordinary explanations of meaning often employ 'ostension' – crudely, pointing. Thus, one may point to Hilary Putnam and utter (5):

(5) **That** is Hilary Putnam

thereby explaining the meaning of *Hilary Putnam*. This is an 'ostensive definition.'

An 'enumerative definition' serves to explain the meaning of an expression by listing at least some elements in the expression's extension. So, for example, (6):

(6) A **Beatle** is Ringo or John or Paul or George

Ostension can be used to facilitate enumeration:

(7) **This** (pointing to Ringo) and **that** (pointing to Paul) are **living Beatles**

Often, enumerative definitions give only partial lists and include an 'and so forth' clause:

(8) A **philosopher** is Hilary Putnam, or W. V. Quine, or Rudolf Carnap, or anything relevantly like those

In (8), since there are indefinitely many ways of continuing the list, we rely on our audience's capacities, in particular the continuations they find salient. In order to reduce reliance, we can give additional information concerning the similarities we expect our audience to track:

(9) A **philosopher** is Hilary Putnam, or W. V. Quine, or Rudolf Carnap, or other things similar to those with respect to academic expertise

An important range of cases involves ostensive enumeration and direction. So, for example,

(10) A sample is **water** iff it is the same liquid as that (pointing to a sample)

According to (10), whether a novel sample counts as water depends on the general requirements on sameness of liquid and on the nature of the original sample. As Hilary Putnam argued, both "... may take an indeterminate amount of scientific investigation to determine" (Putnam, 1975c: 225). Arguably, something close is true of definitions of many ordinary expressions, especially those that employ examples. Development of definitions that are less reliant on context for their functioning than ordinary definitions may require detailed investigation of elements in the circumstances of definition.

Uses Again

The utility of definition depends on how widely it is applicable. There are grounds for pessimism. One negative argument is that, in order for a definition to secure uptake, the *definiens* must be understood. Hence, some basic range of expressions must be understood in advance of any definition, and they will therefore be undefinable. If some expressions can be so understood, it becomes pressing to show that some others cannot. Another negative line is that the role of context in the explanation of meaning establishes that exposure to definitions is, in general, not necessary or sufficient to secure audience understanding. Exposure is insufficient, not only because of the role of context in enabling an audience to utilize a definition to fix on a meaning, but also because elements in the context can play a role in fixing a meaning incompatible with the explicit dictates of the definition. The role of context makes definitions only defeasible guides to meaning. The use of examples above in explaining the varieties of definition makes possible the development – or defeat – of the proffered general characterizations of that variety. Exposure to definitions is unnecessary for a related reason: just as contextual elements can defeat definitions, so they can enable understanding in the absence of definitions.

From the current perspective, these points do not apply to the activity of definition. Since we acquire knowledge of the meanings of many (if not all) of our expressions on the basis of others' explanations – understood to include their uses of expressions in contexts – many expressions with which we are competent are thereby definable. What the points suggest is that meaning can fail to supervene on information acquired just through understanding the *definiens*. (Supervenience of a set of properties *Q* on a set of properties *P* requires that no two possible worlds – or portions of a possible world – can differ in the distribution of *Q*-properties without differing in the distribution of *P*-properties.) But failure of meaning to supervene on the information carried by definitions is perfectly compatible with that information playing a role in sustaining knowledge of meaning. So the two lines of argument canvassed above indicate, at most, that not every ordinary definition will exhibit the degree of freedom from context shown by definitions in logic. The importance of this (potential) result derives from the extent to which philosophers have aimed to offer definitions of key terms – e.g., knowledge, causation, or truth – in a (comparatively) context-free way. One of the major themes of late 20th century philosophy has been that the aim is inappropriate (Burge, 1993; Putnam, 1975a, 1975b, 1975c; Travis, 1989; Wittgenstein, 1953. Related issues arose from Quine's

critique of the view that definitions have distinctive epistemic status: Quine, 1936, 1951, 1963).

See also: Analytic Philosophy; Analytic/Synthetic, Necessary/Contingent, and *a Priori/a Posteriori*. Distinction; Externalism about Content.

Bibliography

- Belnap N D (1993). 'On rigorous definitions.' *Philosophical Studies* 72, 115–146.
- Burge T (1993). 'Concepts, definitions, and meaning.' *Metaphilosophy* 24, 309–325.
- Carnap R (1928). *Der logische Aufbau der Welt*, Berlin: Weltkries; 2nd edn., (1961). Berlin: Felix Meiner; trans. George R (1969). *The logical structure of the world*, Berkeley, CA: University of California Press.
- Carnap R (1947). *Meaning and necessity: a study in semantics and modal logic*. Chicago, IL: University of Chicago Press; 2nd, enlarged edn., (1956).
- Coffa J A (1991). *The semantic tradition from Kant to Carnap: to the Vienna Station*. Cambridge: Cambridge University Press.
- Evans G (1979). 'Reference and contingency.' *The Monist* 62. In Evans G (ed.) (1985) *Collected papers*. Oxford: Clarendon Press. 178–213.
- Fetzer J H, Shatz D & Schlesinger G N (eds.) (1991). *Definitions and definability: philosophical perspectives*. Dordrecht: Kluwer Academic.
- Fodor J A (1998). *Concepts: where cognitive science went wrong*. Oxford: Clarendon Press.
- Kripke S (1980). *Naming and necessity*. Oxford: Blackwell.
- Putnam H (ed.) (1975). *Mind, language and reality: philosophical papers* (vol. II). Cambridge: Cambridge University Press.
- Putnam H (1975a). 'The analytic and the synthetic.' In Putnam (ed.) 33–69.
- Putnam H (1975b). 'Explanation and reference.' In Putnam (ed.) 196–214.
- Putnam H (1975c). 'The meaning of "meaning."' In Putnam (ed.) 215–271.
- Quine W V (1936). 'Truth by convention.' In Lee O H (ed.) *Philosophical essays for A. N. Whitehead*. New York: Longmans; repr. in his 1976, *The ways of paradox* (revised and enlarged edn.), Cambridge, MA: Harvard University Press. 77–106.
- Quine W V (1951). 'Two dogmas of empiricism.' *The Philosophical Review* 60. 20–43; In Quine W V (ed.) (1961) *From a logical point of view*, (2nd edn.). Cambridge, MA: Harvard University Press. 20–47.
- Quine W V (1963). 'Carnap on logical truth.' In Schilpp P A (ed.) *The philosophy of Rudolf Carnap*. LaSalle, IL: Open Court; repr. in his *Ways of paradox* 107–132.
- Robinson R (1950). *Definition*. Oxford: Oxford University Press.
- Russell B (1905). 'On denoting.' *Mind* 14, 479–493; repr. in Marsh R (ed.) (1956) *Logic and knowledge*. London: Allen & Unwin: 41–56.
- Russell B (1927). *The analysis of matter*. London: Kegan Paul.
- Sager J C (2000). *Essays on definition*. Amsterdam: J. Benjamins.
- Soames S (2003). *Philosophical analysis in the twentieth century* (vol. 2). Princeton, NJ: Princeton University Press.
- Suppes P (1957). *Introduction to logic*. Princeton, NJ: Van Nostrand.
- Travis C (1989). *The uses of sense*. Oxford: Oxford University Press.
- Whitehead A N & Russell B (1910). *Principia mathematica* (vol. 1). Cambridge: Cambridge University Press; 2nd edn., (1925).
- Wittgenstein L (1953). *Philosophical investigations*. New York: Macmillan.

Deflationism

G Marti, ICREA & Universitat de Barcelona, Barcelona, Spain

© 2006 Elsevier Ltd. All rights reserved.

The label 'deflationary' is applied to different theories of and approaches to philosophical issues nowadays, and there is a bit of confusion as regards what it is that makes a theory deflationist, partly because it is not clear what all so-called deflationary theories have in common. At the risk of oversimplifying, let us say that a philosophical theory of X is deflationary just in case it does not accord to X a fundamental, or underlying, nature amenable to philosophical analysis.

A deflationary theory of X does not deny the existence of X or of things that are X; it doesn't have to deny that either X or the concept of X plays a role in philosophical inquiry about other phenomena; it does not even have to deny that there may be an explanation of why something is X. The important point to remember is that a deflationary explanation does not constitute a philosophical analysis, nor does it presuppose that there is a uniform reason for being X for all things that are X, i.e., a general explanation of what makes something X – for that would presuppose that X has a nature. Here we will focus on two forms of deflationism that are of special interest to the Philosophy of Language: deflationism about truth and about meaning.

Deflationary conceptions of truth have been espoused under different guises by philosophers such as Gottlob Frege, Frank Ramsey, A. J. Ayer, Peter Strawson, and W. V. O. Quine. Main contemporary proponents include Hartry Field and Paul Horwich.

Ramsey (1927) and Ayer (1936) have both stressed in different ways Frege's remark that "nothing is added to the thought by my ascribing to it the property of truth" (1918), p. 88. To say of a sentence *p*, or of the proposition it expresses, that it is true does not add anything to saying *p* itself. In a similar vein, Quine remarked that "by calling the sentence ('Snow is white') true, we call snow white. The truth predicate is a device of disquotation" (1986), p. 12. This does not mean that the predicate 'is true' is useless, nor that it can be eliminated. Indeed, in some sense a deflationist may rest content with the claim that 'is true' expresses a property: contemporary deflationists accept that 'is true' functions grammatically like a predicate, so like every other predicate it can be said to apply a property to a subject. Nevertheless, it makes no sense to think of truth as the property that all true sentences, or true propositions, have in common, for what makes 'Snow is white' true is something having to do with snow and the physical phenomena involved in the reflection and perception of light, whereas what makes 'London is north of Barcelona' true has to do with the cities and their position on the planet; there is no one thing that these two sentences have in common that explains their being both true. This contrasts with the claim that the two sentences have in common their correspondence to facts in the world, as versions of the correspondence theory would have it, and it contrasts also with the claims that the sentences cohere with a system of beliefs or that they both have some sort of pragmatic value, as the coherence and the pragmatic approaches to truth would have. The predicate 'is true' is not eliminable, because it allows us to say things that we could not say without it – for instance, sentences such as 'Everything Tom says is true,' something that can be meaningfully said even when we do not know all the members of the set of things that Tom says. So, 'is true' has a useful application when it is meant to apply to a conjunction of statements not all of whose members are known, or when it is meant to apply to an infinite conjunction of statements. But observe that on such views the function of the predicate 'is true' is purely logical: there is nothing more to the nature of truth than the linguistic function of 'is true.' And as regards the concept of truth, the deflationist proposes that it is completely captured by Tarski's equivalence schema: *p* is true if and only if *p*.

Some contemporary versions of deflationism assume the instances of the schema to be about sentences, so the nature of truth is taken to be entirely

captured by the list of instances: "'Snow is white' is true if and only if snow is white," "'London is north of Barcelona' is true if and only if London is north of Barcelona," . . . This approach, which finds its inspiration in Quine (1986), faces the objection that it makes truth relative to a language. So, what this version of the deflationist project really captures is the notion of 'true-in-English' or 'true-in Spanish,' something that is arguably at odds with the intuitive notion of truth. A different deflationary approach to truth, the minimalist approach championed by Paul Horwich (1990), regards the instances of the T-schema as being about propositions. So the minimalist view is that the nature of truth is entirely captured by instances of the schema: the proposition that *p* is true if and only if *p*. A proposition is what is expressed by an utterance of a sentence in a given context; utterances of, say, 'Snow is white' by an English speaker and of 'La neige est blanche' by a speaker of French express the same proposition. Thus the notion of truth so characterized is not language-relative.

Horwich's minimalist conception of truth has been widely discussed recently. I will not make an attempt here to give a summary of objections and replies. A good state-of-the-art perspective on the discussion can be obtained from Blackburn and Simmons (1999) and an exhaustive presentation of deflationism, its different versions, and the challenges they face can be found in Kirkham (1992). There is, however, a general worry about the deflationist perspective on truth that is worth mentioning. Robust theories of truth such as the correspondence theory take seriously the explanatory relationship between facts, the way things are in some portion of the world, and the truth of sentences that describe those portions of the world. Thus, we assume that snow's being white *explains* why 'snow is white' is true, something that the deflationist also accepts. So, let us then say: "that snow is white explains why 'snow is white' is true," or alternatively "that snow is white explains why the proposition that snow is white is true." The deflationist seems committed to the claim that the latter says nothing over and above "that snow is white explains that snow is white." Perhaps the deflationist may want to answer to this worry by denying that the original intuition, that snow's being white *explains* why 'snow is white' is true, tells the whole story: that it is not snow's being white, but the physical (not philosophical nor semantic) phenomena about the reflection of light and its perception by the human eye that explain why 'show is white' is true.

There is another conception of truth that in a sense does not qualify as deflationary. Yet in another sense it does. This is the approach to truth explored in Cartwright (1987) 'A neglected theory of truth' (an approach for which Cartwright found inspiration

in G. E. Moore), a view according to which truth is a simple, unanalyzable property. True sentences, or propositions, have the property, false ones do not, and that's all that can be said about true and false sentences or propositions. In some sense the view is everything but deflationary, for it accords to truth a nature and a substantial worldly presence. On the other hand, the stance on nonanalyzability goes hand in hand with other deflationary approaches. It is easy to dismiss Cartwright's 'neglected theory of truth' as a non-theory, for it postulates the existence of a property while denying that there is any explanation to be given as regards the conditions for its possession. This sounds akin to claiming that some sentences express necessary truths, while happily declaring that nothing else needs to be said about necessity and contingency. This form of criticism may be blind to an important distinction, one that is important for all forms of deflationism to highlight: it is one thing to reject the need for explanation, and a different thing to deny that there is a **philosophical** explanation of the conditions of possession of a property (or that there is a philosophical explanation of the conditions of application of a concept). Take for instance the concept red and the property of being red. There are lots of things that a physicist can tell us about the conditions for the possession of that property. In this sense, redness is ripe for explanation. Yet, from a philosophical point of view, one may wish to deny that the concept red is amenable to philosophical analysis, and one may regard the property of being red as a basic property that things either have or do not have.

Deflationary accounts of meaning have in common with deflationism about truth the idea that meaning is not some kind of analyzable property that certain signs possess. Use theories of meaning are deflationist, for they postulate that all there is to the meaning of an expression is the use that speakers make of it. Meaning, according to use theories, does not consist in a relation between words and some entity, mental or worldly. As Horwich (1998) put it: "the meaning of a word is constituted from ... the regularities governing our deployment of the sentences in which it appears" (p. 3).

In Horwich's view the meaning of, for instance, 'dog' reduces just to "its use being governed by such-and-such regularity" (p. 6). The appeal to use makes the explanation of the meaning of 'dog' non-semantic. In purely semantic terms all we can say about the meaning of 'dog' is that "dog" means DOG."

As in the theory of truth, there is a conception of meaning, and in particular of the referential link between expressions and the things they name, that is deflationary in the sense that Cartwright's 'neglected

theory' is deflationary. Proponents of Millianism, direct reference theorists, argue that there is no semantic mechanism (descriptive backup or Fregean mode of presentation) connecting a name and its bearer: that reference is direct and unmediated. As was the case with the Moorean theory of truth, it is easy to dismiss this position as a non-theory: if reference cannot be explained or analyzed, if we can say nothing about reference over and above saying that 'Aristotle' refers to Aristotle, it seems that our words refer to pieces of the world by magic. However, what the view really says is that there is no *philosophical* analysis or explanation that will tell us in virtue of what 'Aristotle' refers to Aristotle: the explanation, interesting as it may be, will involve perhaps social, historical, anthropological, physical, and psychological factors.

See also: Direct Reference; Reference: Philosophical Theories; Truth: Theories of in Philosophy; Use Theories of Meaning.

Bibliography

- Ayer A J (1936). *Language, truth and logic*. London: Gollancz.
- Blackburn S & Simmons K (eds.) (1999). *Truth*. Oxford: Oxford University Press.
- Cartwright R (1987). 'A neglected theory of truth.' In Cartwright R (ed.) *Philosophical Essays*. Cambridge, MA: MIT Press. 71–93.
- Field H (1986). 'The deflationary conception of truth.' In MacDonald G & Wright C (eds.) *Fact, science and morality. Essays on A. J. Ayer's Language, truth and logic*. Oxford: Blackwell. 55–117.
- Field H (1994). 'Deflationist views of meaning and content.' Reprinted in Blackburn S & Simmons K (eds.) (1999). 351–391.
- Frege G (1918). 'The thought: a logical inquiry.' Reprinted in Blackburn S & Simmons K (eds.) (1999). 85–105.
- Gupta A (1993). 'A critique of deflationism.' Reprinted in Blackburn S & Simmons K (eds.) (1999). 282–307.
- Horwich P (1990). *Truth*. Oxford: Blackwell.
- Horwich P (1998). *Meaning*. Oxford: Clarendon Press.
- Kirkham R L (1992). *Theories of truth*. Cambridge, MA: MIT Press.
- Quine W V O (1986). *Philosophy of logic*. Cambridge, MA: Harvard University Press. [See esp. chapter 1.]
- Ramsey F P (1927). 'On facts and propositions.' Reprinted in Blackburn S & Simmons K (eds.) (1999). 106–107.
- Strawson P F (1950). 'Truth.' Reprinted in Blackburn S & Simmons K (eds.) (1999). 162–182.
- Tarski A (1944). 'The semantic conception of truth and the foundations of semantics.' Reprinted in Blackburn S & Simmons K (eds.) (1999). 115–143.

Deixis and Anaphora: Pragmatic Approaches

K Green, Sheffield Hallam University, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

'Deixis' is generally understood to be the encoding of the spatiotemporal context and subjective experience of the encoder in an utterance. Terms such as *I*, *here*, *now*, and *this* – the so-called 'pure deictic terms' – are heavily context dependent and represent a kind of cognitive center of orientation for the speaker. What, for instance is *here* for me, may be *there* for you. Clearly such terms pose problems in terms of both reference and meaning, and standard accounts have attempted to find a middle ground between lexical and pragmatic meaning (see Green, 1995). The difference between 'anaphora' and deixis is fairly straightforward again in standard accounts (see Jarvella and Klein, 1982), but an increasing pragmatic emphasis has made the distinction between the two less easy to define. In standard accounts, anaphora is seen as much more of an intralinguistic or intrasentential element. Consider the following sentences:

That man is very tall. He must have trouble buying clothes.

The deictic expression *that man* must be given a pragmatic interpretation, while the pronoun *he* is said to 'refer back' to the foregoing element. But as we shall see, the issue is not so simple.

Anaphora is generally understood to be the process whereby a linguistic element is interpreted derivatively from a foregoing unit – its 'antecedent.' Although it covers a range of expressions that the speaker may use in referring and picking out the intended referent, research has focused almost exclusively on pronominal referring expressions in discourse, for example:

Fred came into the room. He sat down.

The theory of anaphora deals with the relationship between *he*, *Fred*, and the objects that these elements describe or pick out. The most problematic and interesting of anaphoric phenomena are those that are crosssentential or discourse based. In the taxonomy of Hankamer and Sag (1977), the antecedent is not considered to be crucial, and where it is not explicitly stated, the process is known as 'pragmatically controlled anaphora.' However, this is not central to the kind of pragmatic approaches that have recently been explored and indeed is close to what Halliday and Hasan (1986) and Brown and Yule (1983) would call 'exophoric' reference.

Pragmatic approaches to anaphora are part of a larger set of (sometimes conflicting) theories that Breheny (2002) calls 'nondynamic.' Within these

non-dynamic approaches there are two subsections: 'linguistic' and 'pragmatic' approaches, the latter including what has come to be known as (following Cooper, 1979) 'E-type approaches'. A nondynamic linguistic approach to the above example would stipulate that the pronoun in the second sentence is bound to the noun phrase in the first. However, one can still promote a 'linguistic' approach without the idea of crosssentential binding. In this approach, the pronoun serves as a proxy for a definite description and would therefore easily be accommodated within Hankamer and Sag's model. The relationship between antecedent and pronoun is not so much syntactic as paradigmatic; the pronoun deputizes for the noun phrase. Linguistic approaches nevertheless agree that some linguistic rule enables the description to be recovered, whether due to semantic or syntactic considerations (Heim, 1990; Neale, 1990) (*see Semantics–Pragmatics Boundary*).

Pragmatic approaches stress the absence of any generalized linguistic rule that would account for the recovery and interpretation of descriptions in anaphoric contexts. Rather, any principles about the process of recovery must be inferred from more general principles about discourse organization and context. Consider the following example from Breheny (2002):

A man who walked in the park whistled.

Putting aside for the moment whether this is an appropriate sentence for analysis (that is, whether it is pragmatically plausible), two problems are evident here. The first is the problem of uniqueness, a particular difficulty when the antecedent is an indefinite (*a man*). In the model of Evans (1977) there must necessarily be only one man walking in the park; in other words, *a man* cannot be a disguised plural, giving rise to an 'attributive' (after Donnellan, 1978) reading of the phrase *a man who*. What linguistic rule tells us that this is so? The past tense of the verb *whistled* leads us away from any notions of generality as they tend to occur in timeless present contexts (*A man walking in the park whistles*). However, this by no means counts as a linguistic rule and is more readily interpreted through pragmatic means.

Another problem arises with contradictions. Again, consider the following from Breheny (2002):

A: Last night I met a Cabinet minister.
B: She was not a Cabinet minister.

Here, B does not ascribe to the description the property that A thinks it has. The pragmatic approach

would have to rely on the notion of 'implicit content' to interpret this exchange. The proposition expressed by A cannot depend on the actual state of the referent. Clearly, pragmatic approaches can neatly sidestep both the uniqueness and contradiction problems, but they do so at a cost. The way the utterances are interpreted does not seem to rely upon any generalized notions of either discourse organization or contextualization. Rather, there is at most a general assumption about implicit communication, and we have no way of predicting which interpretation is correct or at least the most salient. As Breheny notes, this has led many to look further for some kind of linguistic rule; but there is no reason to suppose that because pragmatic rules are difficult to locate there must be a more formal linguistic answer.

Where a pronoun is apparently bound to an indefinite expression without restriction, certain difficulties arise, as in the following example (the so-called 'donkey anaphora'):

Every farmer who owns a donkey beats it.

The problem for nonpragmatic accounts of anaphora here is that the indefinite (with its *wh*-clause) bound to the pronoun (*every farmer who owns a donkey*) is a quantificational noun phrase that actually has the pronoun in its scope.

Pronouns with definites as antecedents would appear to offer less of a challenge to pragmatic approaches, as in the earlier-cited example of *Fred came into the room. He sat down*. The traditional approach is to regard the pronoun as 'coreferring' to the name *Fred*, thus tying the pronoun variable to the noun phrase constant. But this makes the pronominal reference purely intralinguistic (as an element of 'cohesion') and wholly dependent upon its antecedent. A more pragmatic approach would be to see the pronoun as referring in a different way to that which *Fred* refers to. Just how different and in what way is difficult to specify, but certainly this approach makes pronominal reference of this kind very similar to deictic or exophoric reference; the line between deixis and anaphora is blurred. A problem with the traditional approach, which we might call the 'binding' or 'cohesive' approach, is that in indefinite contexts, what appears to be a straightforward binding of a quantificational expression and a variable (pronoun) does not represent a rule that can be generalized to a meaningful degree. In crosssentential examples such as *I had ten marbles but dropped them. I found nine. It had rolled under the sofa*, because the pronoun is not properly bound to its antecedent, there is no plausible interpretation. However, in the following example from Breheny, *Every boy left school early. He went to the beach*,

the pronoun seems to convert the universal quantifier into an existential one.

There is a considerable body of work devoted to promoting pragmatic approaches to anaphoric interpretations. In general these works share the view that anaphoric reference is interpreted by means of a range of inferential strategies. This is essentially a neo-Gricean approach, as evident in the work, particularly, of Huang (1994), who employs Gricean principles proposed by Levinson (1987).

Much of the debate hinges on the very nature of pronouns and their contexts. Pronouns do not contain what is traditionally thought of as 'descriptive material,' and yet they are used to refer precisely to that material. This has led quite naturally to theories of anaphoric behavior broadly deemed 'substitutional.' Perhaps the most radical attack on these approaches was presented by Jones (1995), who declared the phenomenon of anaphora and its attendant substitution theories a 'hoax.' For Jones, reference is not an intrinsic property of particular language units and therefore not a property that some units have more than others. Broadly pragmatic, Jones's theory sees reference as a contextualized communicative action expressed in and through the properties of an utterance as a whole. Since, for Jones, what is referred to by speakers is outside and beyond the means used to refer to it, there is no possibility of any linguistic constraints on reference. Such a radical view represents the ultimate pragmatic position, but most theorists have been content with a compromise, whereby linguistic elements routinely prompt a number of interpretive strategies. However, such a compromise is rarely satisfactory and often results in a weakened pragmatics, wherein what appears to be a contextual rule is in fact no more than a linguistic one, dressed up in the language of Gricean or neo-Gricean pragmatics.

Furthermore, there seems to be evidence that anaphoric pronouns are not merely interpreted in a pragmatic fug of inferencing, but are, after all, subject to certain constraints. Jones's view therefore throws too much onto context and pragmatics; after all, if there were no constraints or intralinguistic rules, we would not be able to make the right kinds of inferences.

As Heim states:

... there are data which seem to point to the existence of tighter and somehow more 'syntactic' limitations on the range of reading that actually emerge (p. 165).

Deixis is much more easily subsumed under a pragmatic theory. Traditional accounts make a distinction between the indexical and symbolic meanings of deictic terms. The symbolic meaning of a deictic term might be said to be its semantic aspect, while the indexical meaning is its pragmatic aspect.

For example, the expression *here* might be said to have a symbolic aspect roughly glossed as proximity to the speaker, while the indexical aspect would be the precise location to which *here* was referring. The terms ‘indexical’ and ‘symbolic,’ then, have much in common with the Fregean concepts of ‘reference’ and ‘sense’ (see **Sense and Reference: Philosophical Aspects; Reference: Philosophical Theories**). But the semantic aspect is so vague in certain contexts that many have felt that it cannot really be said to have any bearing on interpretation. Nunberg (1993) in particular has shown that, just as the indexical meanings of deictic terms change according to the contexts in which they occur, so too do the symbolic meanings. This breaking down of traditional binarism has had important implications for pragmatic theory. Emphasis has shifted from the meaning and reference of terms in possible contexts to consideration of the cognitive methods that addressees employ in the interpretation of utterances. Pragmatic approaches have in general attempted to blur the line between deixis and anaphora, but the tendency to see one element (deixis) as essentially exophoric and the other (anaphora) as intralinguistic remains.

See also: Reference: Philosophical Theories; Semantics–Pragmatics Boundary; Sense and Reference: Philosophical Aspects; Syntax–Semantics Interface.

Bibliography

Breheny R (2002). ‘Nondynamic analyses of anaphoric pronouns: do things look better in 2-D?’ In Katz G, Reinhart S & Reuter P (eds.) *Proceedings of Sinn and Bedeuting* 6. Osnabruck: Institute of Cognitive Sciences.

Brown K & Yule G (1983). *Discourse analysis*. Cambridge: Cambridge University Press.

Cooper R (1979). ‘The interpretation of pronouns.’ In Heny F & Schnelle H (eds.) *Syntax and semantics 10: Selections from the third Groningen round table*. New York: Academic Press. 61–92.

Donnellan K (1978). ‘Speaker references, descriptions and anaphora.’ In Cole P (ed.) *Syntax and semantics 9: Pragmatics*. New York: Academic Press. 47–68.

Evans G (1977). ‘Pronouns, quantifiers and relative clauses.’ *Canadian Journal of Philosophy* 7, 467–536.

Green K (ed.) (1995). *New essays in deixis*. Amsterdam: Rodopi.

Grice H P (1975). ‘Logic and conversation.’ In Cole P & Morgan J C (eds.) *Syntax and semantics 3: Speech acts*. New York: Academic Press. 4–58.

Halliday M A K & Hasan R (1986). *Cohesion in English*. London: Longman.

Hankamer J & Sag I (1977). ‘Syntactically versus pragmatically controlled anaphora.’ In Fasold R W & Shuy R (eds.) *Studies in language variation*. Washington, D.C.: Georgetown University Press. 120–135.

Heim I (1990). ‘E-type pronouns and donkey anaphora.’ *Linguistics and Philosophy* 13(2), 137–177.

Huang Y (1994). *The syntax and pragmatics of anaphora*. Cambridge: Cambridge University Press.

Jarvella R J & Klein W (eds.) (1982). *Speech, place and action*. Chichester: John Wiley.

Jones P (1995). ‘Philosophical and theoretical issues in the study of deixis: a critique of the standard account.’ In Green (ed.). 27–48.

Levinson S (1987). ‘Pragmatics and the grammar of anaphora.’ *Journal of Linguistics* 23, 379–434.

Neale S (1990). *Descriptions*. Cambridge, MA: MIT Press.

Nunberg G (1993). ‘Indexicality and deixis.’ *Linguistics and Philosophy* 16, 1–43.

Stalnaker R (1999). *Context and content*. Oxford: Oxford University Press.

Description and Prescription

G Nelson, University College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

In linguistics, a ‘descriptive’ approach involves, as far as possible, the objective and nonjudgmental description of language use. In contrast, a ‘prescriptive’ approach involves laying down, or prescribing, rules for correct use. The dichotomy between the two approaches most commonly surfaces in the area of grammar, though it is also relevant in other areas, such as vocabulary and phonology. The contrast between description and prescription can be illustrated in the following example: A descriptive linguist

(‘descriptivist’) will study the use of *ain’t* (as in, for instance, *I ain’t ready yet*) in terms of regional and social variation, in terms of the distribution of the usage in formal and informal contexts, and will perhaps also study the history and development of the expression. A prescriptive linguist (‘prescriptivist’) will judge the expression as constituting either ‘good’ or ‘bad’ grammar, and will condemn its use if it is seen to fail to measure up to the norms of the standard language. In the broadest terms, then, the descriptivist aims to record language, while the prescriptivist aims to shape language. However, as will be shown, the two approaches are not quite as diametrically opposed as it may appear at

first sight. The two approaches exist on the same spectrum.

For the vast majority of people, the prescriptive approach is by far the most familiar. Indeed, for most nonspecialists, the term ‘grammar’ is virtually synonymous with a set of prescriptive rules that distinguish between ‘good grammar’ and ‘bad grammar’, between the standard variety of a language and the nonstandard varieties, and between grammatical and ungrammatical expressions. In the study of the English language, prescription has had a long history, and one that continues to the present day. Indeed, almost all early grammars of English were largely prescriptive and have contributed an array of prescriptive rules with which most native speakers are very familiar. These include the following familiar axioms:

- Do not split an infinitive (e.g., *to boldly go*).
- Do not start a sentence with *and*.
- Do not end a sentence with a preposition.

Descriptivists would strongly disagree with these rules, since they have no basis in English grammar. These and other rules are the legacy of a long tradition of writing about English from a prescriptive or ‘normative’ point of view. As early as the Renaissance, many writers were concerned about foreign borrowings into English. Writing in 1561, Sir John Cheke was particularly forthright in his views: “I am of this opinion that our tung should be written cleane and pure, unmixt and unmangeled with borrowing of other tungen . . .” (quoted in Hughes, 2000: 155).

Underlying comments such as Cheke’s is a notion of ‘plain’ or ‘pure’ English, which most Renaissance commentators associated with native, Anglo-Saxon vocabulary (Jones, 1953). This may be seen as the beginning of the prescriptive tradition in English, in which writers attempted to shape the language being used by identifying – however informally – a standard to which users should aspire. For Renaissance writers, this standard corresponded to ‘purity’ or ‘propriety’ of expression.

The prescriptive approach to English reached its height during the 18th century, to such an extent that the century became known in the history of English as the ‘age of correctness’ (Leonard, 1929). The approach is best exemplified in Jonathan Swift’s *Proposal for correcting, improving, and ascertaining the English tongue* (1712). Swift proposed the establishment of an Academy for English, along the lines of the Académie Française, which would act as a final arbiter in linguistic matters. Together with many of his contemporaries, Swift felt that arbitration was needed, especially in matters of grammar and vocabulary:

Besides the Grammar-part, wherein we are allowed to be very defective, they will observe many gross improprieties, which however authorised by Practice, and grown familiar, ought to be discarded. They will find many words that deserve to be utterly thrown out of our Language, many more to be corrected, and perhaps not a few, long since antiquated, which ought to be restored. [Jonathan Swift, *Proposal for correcting, improving, and ascertaining the English tongue* (1712)]

As Swift put it, his proposal was designed to find a way of “ascertaining and fixing our Language for ever”, an aspiration that in varying degrees remains central to the prescriptive approach. For a variety of reasons – mostly political – Swift’s proposal was never acted on, and English has never had a central regulating body like the Académie Française. Instead, it has had, and continues to have, a series of more or less self-appointed authorities on the language. In the 18th century, Robert Lowth became a highly influential authority on English grammar, following the publication in 1762 of his *A short introduction to English grammar*. In the preface to his grammar, Lowth set out his prescriptive approach unambiguously:

The principal design of a Grammar of any Language is to teach us to express ourselves with propriety in that Language; and to enable us to judge of every phrase and form of construction, whether it be right or not. The plain way of doing this is to lay down rules, and to illustrate them by examples. [Lowth, 1762: x]

This is probably the clearest and most emphatic statement of the prescriptivist approach, at least insofar as it pertains to grammar. Prescriptivism is primarily didactic and regulatory. Lowth’s work is replete with examples of ‘false’ grammar, taken from literature, which he then corrects in the most unambiguous terms. Thus, for example, he takes Locke to task for his ‘improper’ use of *who*: “‘We are still much at a loss, *who* civil power belongs to’. Locke. It ought to be *whom*” (Lowth, 1762: 127). Lowth’s work did, in fact, contain a great deal of grammatical description, based on a model of Latin grammar, but it is his prescriptivism – his emphasis on identifying and expunging errors – that is his lasting legacy. His work became a standard and highly respected authority on English grammar for many decades.

In vocabulary and word meanings, the same function was performed by Samuel Johnson’s *A dictionary of the English language* (1755). In Johnson’s work, however, we see the tension between prescriptivism and descriptivism being played out in a highly perceptive way. Johnson began his work in 1747, when he published his *Plan of a dictionary*. In that work, he set out his aims and objectives, and comes across as following firmly in the prescriptive tradition. To a

large extent, he saw his role as a ‘defender’ or ‘gate-keeper’ of the language, regulating the vocabulary of English in the same way that Lowth sought to regulate its grammar. Of “barbarous or impure” words, he wrote, these “may be branded with some note of infamy, as they are to be carefully eradicated wherever they are found” (Lynch, 2002: 577). Johnson used an extensive range of usage labels, including ‘bad’, ‘barbarous’, ‘low’, ‘ludicrous’, and ‘rustick’. His prescriptivism can be seen clearly throughout his dictionary, including in the following entries:

To bamboozle v.a. A cant word not used in pure or grave writings.

To dumbfound v.a. A low phrase.

Shabby adj. A word that has crept into conversation and low writing; but ought not to be admitted into the language.

For modern readers, there is a curious tension between Johnson’s recording the use of certain words, only to prescribe that they should not be used. In these examples, and in many others, his perception of the lexicographer’s role as gatekeeper is quite clear. While on the one hand his role is to record the language, he also sees it as his duty to regulate the language, and to provide instruction to his readers. He may, in fact, have been aware of this tension himself, since in the *Plan* he alludes to his dual role: “[I]n lexicography, as in other arts, naked science is too delicate for the purposes of life. The value of a work must be estimated by its use: it is not enough that a dictionary delights the critic, unless at the same time it instructs the learner” (Lynch, 2002: 566).

Striking the balance between ‘naked science’ and didacticism remains a central problem in linguistics. In Johnson’s case, and certainly in his *Plan*, the balance is strongly in favor of the didactic, prescriptive end of the spectrum. His stated objective was to produce “a dictionary in which the pronunciation of our language may be fixed, and its attainment facilitated; by which its purity may be preserved, its use ascertained, and its duration lengthened”. His description of the lexicographer’s work in terms of military invasion and conquest constitutes one of the strongest statements of the prescriptivist position:

When I survey the Plan which I have laid before you, I cannot . . . but confess, that I am frighted at its extent, and, like the soldiers of Caesar, look on Britain as a new world, which it is almost madness to invade. But I hope, that though I should not complete the conquest, I shall at least discover the coast, civilize part of the inhabitants, and make it easy for some other adventurer to proceed further, to reduce them wholly to subjection, and settle them under laws. [Lynch, 2002: 579]

In the process of compiling the *Dictionary*, however, Johnson’s view shifted somewhat toward a more descriptive stance. A comparison between the *Plan* of 1747 and the *Preface* of 1755 is instructive in this regard, since it rehearses many of the issues that still obtain in the description/prescription debate. At one level, Johnson simply discovered the enormity of the task he had originally set himself, and realized that it was impossible to achieve. But at a deeper level, he came to recognize two important aspects of language. The first of these is the inevitability of change:

Total and sudden transformations of a language seldom happen; conquests and migrations are now very rare: but there are other causes of change, which, though slow in their operation, and invisible in their progress, are perhaps as much superiour to human resistance, as the revolutions of the sky, or intumescence of the tide. [Lynch, 2002: 41]

From this starting point, Johnson developed the idea of the inevitability of change in language, and specifically the notion of necessary innovation:

The language most likely to continue without alteration, would be that of a nation raised a little, and but a little, above barbarity, secluded from strangers, and totally employed in procuring the conveniences of life But no such constancy can be expected in a people polished by arts When the mind is unchained from necessity, it will range after convenience; when it is at large in the fields of speculation, it will shift opinions; as any custom is disused, the words that expressed it must perish with it; as any opinion grows popular, it will innovate speech in the same proportion as it alters practice. [Lynch, 2002: 41]

Johnson also came to recognize the importance of ‘general agreement’ among users of a language, as he put it, as an arbiter in linguistic matters. This led to his questioning the whole notion of authority in language, as it had been perceived by Swift and others. On a practical level, he observed that the academies of the kind Swift had proposed had been largely ineffective. They had been set up “to guard the avenues of their languages, to retain fugitives, and repulse intruders; but their vigilance and activity have hitherto been in vain; sounds are too volatile and subtle for legal restraints; to enchain syllables, and to lash the wind, are equally the undertakings of pride” (Lynch, 2002: 7).

As Johnson came to perceive it, the ultimate authority in language is invested in those who use it, or in ‘general agreement’ among the speech community. Thus he answered Swift’s concerns about obsolete words in the following way:

Swift, in his petty treatise on the English language, allows that new words must sometimes be introduced, but proposes that none should be suffered to become obsolete. But what makes a word obsolete, more than general agreement to forbear it? [Lynch, 2002: 42]

Though this was highly perceptive for its time, we should not equate Johnson's 'general agreement' with any modern concept of usage. For the most part, he invested authority in language in those literary writers that he himself considered to be the best – namely, English authors, from Philip Sidney to the Restoration, “whose works I regard as *the wells of English undefiled*” (Lynch, 2002: 36). Just as it was in the Renaissance, the notion of ‘purity’ as a linguistic ideal was still strong in Johnson's thinking about language. Perhaps the most modern expression of the authority of usage, in Johnson's own time, was made by Joseph Priestley, in *The rudiments of English grammar* (1761):

We need make no doubt but that the best forms of speech will, in time, establish themselves by their own superior excellence: and, in all controversies, it is better to wait the decisions of Time, which are slow and sure, than to those of Synods, which are often hasty and injudicious. [Priestley, 1761: vii]

Between Johnson's time and our own, fundamental changes have taken place in our perceptions of language, and especially in our understanding of language change. Today, the descriptive approach is predominant among professional linguists, many of whom often use the word ‘prescriptive’ pejoratively. However, it would be naïve to think that prescriptivism has been displaced in modern linguistics. On the contrary, it is alive and well, most notably in public perceptions of language. Indeed, the view of grammar and its didactic role that works such as Lowth's expounded is still widely held today. The reason for this is not difficult to discern: a great many people look to grammars to provide guidance, whether on general grammatical points, on writing ‘good’ English, or on matters of disputed usage. The enduring popularity of usage books is testimony to this. In the 20th century, H. W. Fowler's *Modern English usage* has remained in print, in various editions, since 1926. More recently, the phenomenal commercial success in the United Kingdom of Lynn Truss's *Eats, shoots and leaves: the zero-tolerance guide to punctuation* (2003) shows that the market for prescriptive guides is as strong as ever. In almost every case, these prescriptive guides are written by people who are nonlinguists, but who care deeply about linguistic standards. They frequently see their role as defending the language against unnecessary or unwanted change, and sometimes contrast their work with that of professional linguists. The British novelist Kingsley Amis was another in this long tradition of defenders of English, as he made clear in *The King's English: a guide to modern usage* (1997: xii): “I am sustained by reflecting that the defence of the language is too large a matter to be left to the properly

qualified, and if I make mistakes, well, so do they. . .”. Here, there is an implied criticism of professional linguists, who are perceived to have failed in some way in their responsibility to defend the language.

Most professional linguists today make a very clear distinction between prescriptive rules (which generally deal with matters of disputed usage) and descriptive rules (which generally describe matters of undisputed usage, such as subject–verb agreement). In placing the emphasis on description, they would claim that their role is to describe language in actual use, and not to prescribe how it should be used. In the field of grammar, for instance, the urge toward description has been greatly intensified in recent years through the development of corpus linguistics – that is, the use of a large, computerized database (corpus) of samples of language as the basis of study. For example, the *Longman grammar of spoken and written English* (Biber *et al.*, 1999) is based on a corpus of 40 million words of contemporary English, taken from a wide range of spoken and written sources. The *Oxford English grammar* (Greenbaum, 1996) is likewise based on a corpus of 1 million words of contemporary British English, and includes many illustrative citations drawn from sources such as private conversations, telephone calls, lectures, and business letters. In theory, at least, corpus-based grammars such as these are more ‘objective’ than, say, Lowth's, since the information they contain is not based exclusively on the grammarian's own intuitions or personal preferences, and is not shaped by his/her own subjective judgments on grammaticality and acceptability. Corpus-based grammars also recognize variation within language to a much greater degree than hitherto. By examining both speech and writing, and both formal and informal usage, the corpus-based grammars offer a more representative account of contemporary language. So as well as describing a grammatical feature, they also discuss its distribution, whether in terms of register, geography, or other sociolinguistic variables. In contemporary, descriptive grammar, the ultimate objective is to describe ‘common usage’, in a largely nonjudgmental fashion.

For some members of the general public, however, the rise of descriptive grammars, whether corpus-based or not, is often seen as an unwelcome development. From a prescriptive point of view, simply describing how people use a language is not good enough. The nonjudgmental approach adopted by the descriptivists is seen by many people as leading to the view that, in matters of usage, ‘whatever is, is right’. Parents and teachers, in particular, may view this kind of anything-goes approach as, at best, unhelpful, and, at worst, irresponsible. This is often because it appears to many people that grammarians who refuse to be

judgmental are denying the existence of standards altogether, and are refusing to make the crucial distinction between ‘good’ uses of language and ‘bad’ uses of language. For the prescriptivists, the grammarian’s role in particular should be to make this distinction clear, just as it was for Lowth, so that in learning grammar, their children will learn, specifically, *good* grammar, and thereby develop their literacy skills.

The descriptivists and the prescriptivists have been poles apart for much of the 20th century. The distance between the two approaches is well illustrated in a recent (prescriptive) publication about standards of English, entitled *Lost for words: the mangling and manipulation of the English language* (2004), by the British Broadcasting Corporation journalist, John Humphrys. Among other recent developments in English, Humphrys condemns the loss of the apostrophe in the word *let’s*, in cases in which it has singular reference only (as in, *Let’s give you a hand*). He takes the professional (descriptive) linguist Jean Aitchison to task for her comments about this, in her book *Language change: progress or decay?* Taking a long-term view of language change, Aitchison described the loss of the apostrophe in *let’s* in terms of grammaticalization, and continued with “*Let’s*, perhaps now better spelled *lets*, is used as a simple exhortation, and is no longer thought of as a verb plus a pronoun” (Aitchison, 2001: 113). Taking a strongly prescriptivist stance, Humphrys (2004: 35) comments: “Can this be true? A respected academic condoning such vandalism?” He goes on to consider whether a sentence such as *I could of danced all night* will also be dismissed in the same way, as something that can be explained in terms of grammaticalization, and is therefore ‘condoned’, on that basis, by the descriptivists. He concludes: “Far from liberating the young, this sort of nonsense is more likely to harm them” (Humphrys, 2004: 36).

Of course, both sides in this debate have legitimate points to make, in their own terms. However, these points seem to be irreconcilable, because the ultimate objectives of each side are quite distinct. The objectives of the prescriptivists are short-term, immediate, and practical. They aim to provide clear-cut guidance for the general user of language. Their approach is motivated by a well-intentioned desire to maintain standards of literacy and to teach young people, in particular, the differences between good uses of language and bad uses of language. Few if any people – whether descriptivists or prescriptivists – could argue with this objective. On the other hand, the approach followed by Aitchison and other descriptivists is also legitimate, in its own terms. The objective here is to explain grammatical change over time, and in the context of broader linguistic processes, but it stops

short of evaluating that change in any way. The two approaches also address quite different audiences: the prescriptivists write for nonspecialists, whereas the descriptivists, in general, write for fellow linguists. As we saw, Johnson identified this dichotomy quite early on: it is the ever-present dichotomy in language studies between ‘usefulness’ and ‘naked science’.

Can anything be done to bridge the gap between these two approaches? The first point to consider is that the descriptive approach is not, in reality, quite as objective or as ‘scientific’ as we might be led to believe. Even descriptive grammarians make implicit value judgments about their material, and they will be influenced by their own intuitions about the language. In selecting a particular sentence, for instance, to illustrate a grammatical point, the grammarian is implicitly labeling it as ‘grammatical’, in contrast with the infinite number of ungrammatical sentences. In generalizing from observed corpus data, native-speaker grammarians, in particular, will often rely on their own judgments of acceptability. In writing the descriptive *Comprehensive grammar of the English language* (Quirk *et al.*, 1985), the authors freely admitted that “in the absence of sufficient evidence our evaluations sometimes relied on our own experience and feelings” (Greenbaum, 1988: 34–35). Indeed, the authors of that grammar also took account of the prescriptive tradition, on the grounds that it has had an impact on the language that people use, especially in the most formal contexts. They recommend that grammarians “should take account of prescriptive grammar in their descriptive grammars, since prescriptive norms affect use...” (Greenbaum, 1988: 34). In descriptive, corpus-based grammars, the selection of samples to include in the corpus may also be influenced by subjective or evaluative judgments. In dictionaries, too, usage labels such as ‘informal’ or ‘formal’ are frequently applied to words. Although these are less evaluative than those used by Johnson, they do provide evidence of the lexicographer making judgments about words on the basis of a perceived standard. The supposed objectivity of the descriptive approach is also called into question when we consider that all description is influenced, to some degree, by language theory, of whatever persuasion. The notion of a ‘theoretically neutral’ description of a language is more an ideal than a reality. In whatever subfield the linguist works, an entirely objective description of a language is rarely if ever achieved.

The second, more crucial point is the charge, implicit or explicit, that in adopting a descriptive approach, linguists are in some way shirking their responsibility to ‘defend’ language, or to maintain standards. Although few if any linguists today would see their role in this light, this is an important

issue because it determines public perceptions of linguistics. In *Good English and the grammarian*, Sidney Greenbaum considers this issue, and proposes that professional grammarians should become actively engaged in public discussions about usage, bringing their professional training to bear on points of disputed usage. Recognizing that the term 'prescriptivism' is often used pejoratively by linguists, he suggests that the term 'language planning' should be used instead: "Let us then say that grammarians have a responsibility to be language planners" (Greenbaum, 1988: 38). According to Greenbaum, language planners have a crucial role to play, particularly in areas in which no national standard has yet been established, such as in the nonnative Englishes that are emerging in some postcolonial African and Asian countries. In general, Greenbaum does not see a radical dichotomy between descriptivists and prescriptivists. Although they may be very different in terms of their views of language, their objectives, and the audiences they address, the two can come together in that the results of academic linguistic research will eventually filter down to pedagogic grammars and textbooks: "Scholarly grammars have a limited readership, but they eventually influence lower-level grammars, textbooks, and other teaching material, for native students and (in particular) for foreign students" (Greenbaum, 1988: 36).

See also: Language as an Object of Study; Linguistics as a Science; Linguistics: Approaches.

Bibliography

Aitchison J (2001). *Language change: progress or decay?* (3rd edn.). Cambridge: Cambridge University Press.

- Amis K (1997). *The King's English: a guide to modern usage*. London: HarperCollins.
- Biber D, Johansson S, Leech G, Conrad S & Finegan E (1999). *The Longman grammar of spoken and written English*. London: Longman.
- Greenbaum S (1988). *Good English and the grammarian*. London: Longman.
- Greenbaum S (1996). *The Oxford English grammar*. Oxford: Clarendon Press.
- Hughes G (2000). *A history of English words*. Oxford: Blackwell.
- Humphrys J (2004). *Lost for words: the mangling and manipulation of the English language*. London: Hodder & Stoughton.
- Jones R F (1953). *The triumph of the English language: a survey of opinions concerning the vernacular from the introduction of printing to the Restoration*. Palo Alto, CA: Stanford University Press.
- Leonard S A (1929). *The doctrine of correctness in English usage, 1700–1800*. London: Madison.
- Lowth R (1762). *A short introduction to English grammar*. London.
- Lynch J (ed.) (2002). *Samuel Johnson's dictionary: selections from the 1755 work that defined the English language*. Delray Beach, FL: Levens Press.
- Priestley J (1761). *The rudiments of English grammar*. London.
- Quirk R, Greenbaum S, Leech G & Svartvik J (1985). *A comprehensive grammar of the English language*. London: Longman.
- Truss L (2003). *Eats, shoots and leaves: the zero-tolerance guide to punctuation*. London: Profile Books.

Relevant Website

<http://www.ucl.ac.uk> – University College London online; source for a free course on grammar (access through the site index, *English, the Internet grammar of*).

Descriptions, Definite and Indefinite: Philosophical Aspects

G Ostertag, Nassau Community College, Garden City, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Definite descriptions in English take one of two forms: as the definite article *the* concatenated with a nominal (e.g., *table*, *husband*, *game*) or as either a possessive adjective (*her*, *my*) or noun phrase (*everyone's*, *John's*) concatenated with a nominal. Thus, *the table*, *her husband*, *everyone's favorite game*, *my cat*, and *John's bicycle* are all definite descriptions. In contrast, indefinite descriptions take a single form:

as the indefinite article *a* (or *an*) concatenated with a nominal. Examples of indefinite descriptions are *a table*, *an employee*, *a thing I haven't mentioned*, *a friend of Mary's*. Although this classification is not perfect – *the friend of an acquaintance*, although intuitively indefinite, comes out as definite – it conforms to usage standard among philosophers, logicians, and linguists.

According to Bertrand Russell, descriptions – both definite and indefinite – are devices of quantification. That is, both *the F is G* and *an F is G* can be interpreted as expressing a relation between the properties F and G. Since Russell's treatment is by far

the most influential approach to descriptions in the philosophical literature, this entry will focus on his views. It begins by briefly reviewing the motivations behind Russell's mature view on descriptions, which stem in part from inadequacies of his earlier approach, and proceeds to a statement of Russell's mature view. Challenges to this view are then considered, as are alternative proposals.

Russell's Theories of Description

Russell's Early Theory of Denoting

Intuitively, an utterance of the singular sentence *I met Tony Blair* expresses a proposition that, among other things, is about Tony Blair. In virtue of what does this relation of 'aboutness' hold? For Russell, *circa The principles of mathematics* (1903), the proposition that I met Tony Blair is about Mr Blair by virtue of containing him as a constituent. This in turn suggests an answer to a related question, namely, What is the contribution of the expression 'Tony Blair' to the aforementioned proposition? Russell identifies the contribution 'Tony Blair' makes to this proposition with the constituent that enables it to be about Tony Blair – namely, the individual Tony Blair himself (*see* **Direct Reference**).

How are we to understand the parallel contribution the syntactically complex denoting phrase *a man* makes to the proposition *I met a man*? Russell's answer was that it contributes, not an individual, but a certain complex concept – what he called a 'denoting concept.' Russell conceived of a denoting concept by analogy with what he called a 'class concept' (roughly, a property or condition that determines a class of entities). Whereas the nominal *man* contributes a class concept to the proposition that I met a man, the complex phrase *a man* contributes a denoting concept. However, as Russell noticed, denoting concepts possess a puzzling combination of features. For one, the relation between the denoting concept and its denotation is, as he later put it, 'logical,' and "not merely linguistic through the phrase" (Russell, 1905: 41). That is to say, the denoting concept denotes what it does because of something intrinsic to the denoting concept itself, not because of any facts attaching to the denoting phrase that expresses it. Second, a denoting concept is an 'aboutness-shifter' (Makin, 2000: 18). Although the denoting concept associated with *a man* is a constituent of the proposition *that I met a man*, the denoting concept is not what this latter proposition is about.

Third, denoting concepts fail to conform to a principle of compositionality, according to which the meaning of a complex expression is a function of its

structure and the meanings of its constituents (*see* **Compositionality: Philosophical Aspects**). Russell was keenly aware of this deficiency. He remarked that *all men* and *all numbers* seem to be analyzable into a concept associated with the determiner *all* and the respective class concepts *men* and *numbers*, continuing:

But it is very difficult to isolate any further element of *all-ness* which both share, unless we take as this element the mere fact that both are concepts of classes. It would seem then, that "all *u*'s" is not validly analyzable into *all* and *u*, and that language, in this case as in some others, is a misleading guide. The same remark will apply to *every*, *any*, *some*, *a*, and *the*. (Russell, 1903: 72–73)

The inability of the theory of denoting concepts to reflect the compositional nature of denoting phrases is a serious defect of the approach. Not only does surface grammar overwhelmingly suggest that *all men* and *all numbers* possess a common semantic feature; in addition, speakers familiar with the nominal *curator* and with the determiner *all* will understand sentences containing *all curators* in subject position (assuming familiarity with the other expressions), even supposing they have never come across that particular combination. An acceptable theory of denoting phrases cannot leave this phenomenon unexplained.

Russell's Mature Theory

Russell (1905) developed an approach to denoting phrases that avoided each of the difficulties noted above. His revision makes use of the doctrine of contextual definition, or 'meaning in use,' dispensing with the idea that denoting phrases can be assigned meanings 'in isolation.' Rather, each of the aforementioned denoting phrases is defined within its sentential context. While the treatment of indefinite descriptions, if not wholly uncontroversial, is straightforward (*an F is G* is defined as *something is both F and G*), the treatment accorded definite descriptions is rather less intuitive: *the F is G* is defined as *something is both uniquely F and G*. This is equivalent to the conjunction of three claims: *something is F*; *at most one thing is F*; *every F is G*. It can be seen that, in this analysis, *the F is G* will be false if either nothing is F, more than one thing is F, or some F is not G.

Definite Descriptions in *Principia mathematica*

Russell's favored expression of the theory of descriptions is in the formal language of *Principia mathematica*, where the theory is rendered as follows:

$$(R1) \quad G(\iota x)Fx =_{\text{def}} \exists x(\forall y(Fy \equiv y = x) \ \& \ Gx)$$

The definiens is the formal analogue of *the F is G*, with the *iota* phrase corresponding to the definite article; the definiendum is the formal analogue of the sentence *something is both uniquely F and G*. As the definition shows, the surface grammar of $G(\iota x)Fx$ is misleading with respect to its logical form. While $(\iota x)Fx$ takes singular term position, its logical role is not that of a singular term (that is to say, unlike a singular term, it does not even purport to refer). Indeed, given the logical law $\tau = \tau$, taking descriptions to be singular terms allows the derivation of $(\iota x) (Fx \ \& \ \neg Fx) = (\iota x) (Fx \ \& \ \neg Fx)$; this in turn allows the further derivation of the absurdity $\exists x \ x = (\iota x) (Fx \ \& \ \neg Fx)$. (Note that logical systems permitting empty singular terms do not license the second inference; see Lambert, 2003.)

Russell's theory, as encapsulated in R1, avoided the difficulties that plagued the doctrine of denoting concepts: their mysterious ability to determine their denotations logically; their disruption of the aboutness-as-constituency doctrine; and their failure to conform to a principle of compositionality. (It also enabled Russell to explain how a sentence containing a non-denoting description can nonetheless be meaningful.) Since the denoting phrase $(\iota x)Fx$ disappears under analysis, and since the analysans introduces no new denoting phrases, the first two difficulties no longer arise. In addition, R1 shows how every context in which $(\iota x)Fx$ occurs can be replaced with a context that is fully compositional (although the fact that R1 puts logical form at variance with surface grammar has led some to question its usefulness in a compositional semantics for English).

Descriptions and Scope

R1 shows how to eliminate sentences containing $(\iota x)Fx$ and replace them with sentences containing only the familiar logical vocabulary of variables, predicate constants, connectives, and quantifiers. But the definition fails to provide a unique replacement strategy for sentences such as $G(\iota x)Fx \supset p$. Both of the following are consistent with our definition:

$$(1a) \ \exists x(\forall y(Fy \equiv y = x) \ \& \ (Gx \supset p))$$

$$(2a) \ \exists x(\forall y(Fy \equiv y = x) \ \& \ Gx) \supset p$$

The former corresponds to the reading in which the scope of $(\iota x)Fx$ is $G(\iota x)Fx \supset p$, and the latter corresponds to the reading in which the scope of $(\iota x)Fx$ is $G(\iota x)Fx$ (see **Scope and Binding: Semantic Aspects**). Russell's contextual definition of $(\iota x)\Phi x$ in *Principia mathematica* employs an awkward but effective device to eliminate such structural ambiguities:

$$(R2) \ [(\iota x)\Phi x] \ \Psi(\iota x)\Phi x =_{\text{df}} \exists x(\forall y(\Phi y \equiv y = x) \ \& \ \Psi x)$$

The scope of $(\iota x)\Phi x$ is determined by the placement of $[(\iota x)\Phi x]$. The general rule is that the scope of an

occurrence of $(\iota x)\Phi x$ is the entire context to which the scope operator $[(\iota x)\Phi x]$ is prefixed. Using Russell's notation, the readings corresponding to (1a) and (2a) are represented as (1b) and (2b):

$$(1b) \ [(\iota x)Fx] \ (G(\iota x)Fx \supset p)$$

$$(2b) \ [(\iota x)Fx] \ G(\iota x)Fx \supset p$$

Recent developments in syntactic theory provide a more natural method of indicating scope (see May, 1985). What Russell referred to as 'denoting phrases' are, in fact, natural language quantifiers. Since such quantifiers are invariably restricted, the most natural way to represent them is not in first-order logic – the language of *Principia mathematica* – in which quantifiers are unrestricted, but rather in a language permitting restricted quantification. In such a language, a quantifier is the result of concatenating a determiner, subscripted with a variable, with a formula (for example, $[\text{some}_x: x \text{ is human}]$). As with unrestricted quantification, the resulting expression can itself combine with a formula to create a sentence. This allows us to express *the F is G* as $[\text{the}_x: Fx] (Gx)$. Instead of the serviceable but unwieldy (1b) and (2b), we get (1c) and (2c), which are far more natural renderings of the respective contexts:

$$(1c) \ [\text{the}_x: Fx] (Gx \supset p)$$

$$(2c) \ [\text{the}_x: Fx] (Gx) \supset p$$

Note that in this interpretation, descriptions, like quantified noun phrases generally, can be assigned meanings in isolation. Just as the quantifier *all kings* denotes the set of properties possessed by all kings $[\lambda P \ \forall x(\text{King}(x) \supset Px)]$ and *some kings* denotes the set of properties possessed by some kings $[\lambda P \ \exists x(\text{King}(x) \ \& \ Px)]$, the definite description *the king* denotes the set of properties possessed by something which is uniquely king $[\lambda P \ \exists x(\forall y(\text{King}(y) \equiv y = x) \ \& \ Px)]$. In this approach, *the king is bald* is true just in case being bald is among the properties the king possesses. (For details see Westertahl, 1986.)

This machinery pays off handsomely, disambiguating contexts involving the interaction of descriptions with modal operators and with verbs of propositional attitude (see **Modal Logic; Propositional Attitude Ascription: Philosophical Aspects**). In addition, it clarifies the relation between surface grammar and logical form, something that R1 leaves obscure.

Responses to Russell's Theory of Definite Descriptions

Strawson's Critique of Russell

For almost half a century, Russell's theory maintained the status of orthodoxy. But in an article published in

Mind in 1950, the Oxford philosopher P. F. Strawson launched an influential attack. The attack focused on Russell's alleged disregard for ordinary usage, a disregard manifested in Russell's assumption that for every well-formed sentence-type *S* there exists a unique proposition expressed by *S*. In fact, Strawson noted, the same sentence can be used on one occasion to assert one proposition, on a different occasion to assert another proposition, and on a third to assert nothing at all. In order to maintain his view, Russell is forced to provide what amounts to a logical guarantee that in even the least propitious situation a sentence would, however implausibly, possess a definite truth-value. To ensure that *the F is G* expresses a proposition come what may, Russell interpreted it as a complex existential claim, rejecting the intuitive classification of *the F* as a singular term. Strawson claimed that this analysis is contradicted by common usage. If Russell were correct, then the proper response to *The present king of France bald. Is this true?* would be a firm *no*. In fact, the proper response would address the belief betrayed by the utterance, that is, that France is at present a monarchy.

Strawson claimed that to understand the linguistic meaning of an abstract sentence-type *S* requires mastery of the 'rules of use' associated with *S*. In the case of *the present King of France is wise*, the rules require that it be used only when France is a monarchy. To use this sentence seriously and literally is, among other things, to present oneself as believing that the relevant conditions are fulfilled. If, in fact, the belief is false, then the utterance fails to express a proposition – it is something about which 'the question of truth does not arise.'

Yet Strawson's own theory is open to the following challenges. (1) It is not at all clear that utterances containing vacuous descriptions are devoid of propositional content. A contemporary utterance of *yesterday Mick Jagger met the King of France* is, *contra* Strawson, intuitively false. (2) The proposal fails to apply in any obvious way to relational descriptions bound by a higher quantifier. The description *each girl's mother* in (3a) is properly unpacked as *the mother of x*, with the variable bound by the quantifier *each girl*. This is made explicit in (3b):

(3a) Each girl's mother attended.

(3b) $\forall x(Gx \supset A(ry)Myx)$

How such descriptions can be said to refer is a mystery. While there are formal responses to this worry, they are far from being intuitively satisfying (Evans, 1982: 51–57).

Finally, it should be noted that Strawson partly misrepresented Russell's actual view, or at least provided an unnecessarily inflexible interpretation

of it. For example, Russell could happily incorporate the distinction between sentence and utterance that Strawson accused him of overlooking. Indeed, Russell himself provided the context-sensitive expression *my only son* as an example of a description, indicating an awareness that different tokens of the same sentence may express different propositions. This reveals something Strawson failed to recognize: that the distinction has no direct bearing on the question of the logical form of description sentences. Contemporary Russellians, such as Stephen Neale (1990), accept the distinction, seeing in it no fundamental challenge to Russell's theory.

The Ambiguity Thesis

Keith Donnellan (1966) described a phenomenon that neither Russell nor Strawson had considered: that *the F* can be used to refer to a nonF. Donnellan gave as an example an utterance of *Smith's murderer is insane*, made in court on observing the bizarre behavior of a man (call him Jones) accused of murdering Smith. Even if the description 'fits' someone else, Donnellan claims, the utterance clearly refers to Jones, and not the actual murderer, whoever that might be.

But as Saul Kripke (1977) observed, appeals to usage alone cannot contradict Russell's theory. One can use a sentence to communicate a proposition that departs from its literal meaning, so data about usage are in themselves insufficient to mount a successful challenge to Russell (see **Referential versus Attributive**). What is needed is a decisive intuition that, in the imagined utterance, the speaker literally states that Jones is insane. Such an intuition would strongly favor the thesis that definite descriptions are ambiguous between a nonRussellian, 'referential' use and a Russellian, 'attributive' use. Yet few theorists are ultimately willing to commit to the view that, in the imagined utterance, the speaker has literally stated that Jones is insane.

A related argument, due initially to Strawson, provides a much stronger challenge. Utterances of description sentences typically exhibit incompleteness. That is, while the nominal is true of a plurality of objects, the speaker intends nonetheless to speak truly. For example, a librarian issues a polite reminder to a patron by uttering *the book is overdue*. For Russell, this utterance is true just in case there is exactly one book, and any book is overdue. This seems to get things wrong, as the librarian is clearly not intending to say – and is not taken to be saying – something that is, as things are, blatantly false.

Though some have argued that, intuitions notwithstanding, the utterance is, strictly speaking, false, few have found this idea appealing (but see Bach, 1988).

Russellians have responded by claiming that the context of utterance can be relied on in such cases to provide the missing completing material. Thus, what the librarian intends to convey is perhaps that *the book you borrowed last month is overdue* or that *the book you just mentioned is overdue*, and so on. Of course, the suggestion leads immediately to a worry. Typically, when a speaker utters a sentence containing a contextually incomplete description, there will be a multiplicity of completing properties to choose from. The question, Which completion is the one that both the speaker intended to convey and the hearer took the speaker to have intended? can receive no definite answer. And yet, such utterances are not typically accompanied by any indeterminacy or uncertainty as to what was said.

Incomplete descriptions are used almost invariably to refer to a contextually definite object – in our example, to a particular book. In such cases, the speaker succeeds in communicating a proposition containing (or somehow about) the entity in question. Given that such sentences are typically used to communicate just such ‘referential’ propositions, Howard Wettstein (1981) has suggested that in such contexts, a speaker literally asserts the relevant referential proposition. This hypothesis is, after all, consistent with the fact that the librarian’s utterance was perfectly determinate, whereas the Russellian hypothesis, sketched above, is not.

Of course, there remains the case of non-referentially-used incomplete descriptions: for example, *the murderer is insane*, uttered at the scene of a brutal murder and without knowledge of the murderer’s identity. Wettstein’s suggestion is useless here. After all, such uses cannot be supposed to express referential propositions, since there is no referent. But to suppose, with Wettstein, that context provides completing information in such cases is no more plausible here than in the referential case.

Responses to Russell’s Theory of Indefinite Descriptions

Referential Uses of Indefinite Descriptions

As we have seen, Russell defines *an F is G* in terms of the existential quantification, *something is both F and G*. As with definite descriptions, indefinite descriptions are often used referentially. Moreover, the relevant data seem open to the same response – that facts about usage cannot, by themselves, allow one to draw conclusions about literal meaning. But, as in the previous cases, matters are not so simple. Consider (4):

- (4) There’s a man at the door. He’s selling linguistics encyclopedias.

If Russell is correct, then the first sentence asserts that at least one man is at the door. But how then are we to understand the second sentence? After all, intuition suggests overwhelmingly that the pronoun refers to the individual introduced by the indefinite ‘a man.’ Doesn’t this force us to conclude that the indefinite is likewise referential? Not necessarily: as Lewis (1979: 243) noted, “What I said was an existential quantification; hence, strictly speaking, it involves no reference to any particular man. Nevertheless it raises the salience of the man that made me say it.” And this fact allows a subsequent pronoun to make literal reference to the man.

This seems to settle the question in favor of Russell. But an example due to Michael Devitt raises a further difficulty for his theory:

Several of us see a strange man in a red baseball cap lurking about the philosophy office. Later we discover that the Encyclopedia is missing. We suspect that man of stealing it. I go home and report our suspicions to my wife: “A man in a red baseball cap stole the Encyclopedia.” Suppose that our suspicions of the man we saw are wrong but, “by chance,” another man in a red baseball hat, never spotted by any of us, stole the Encyclopedia. (Devitt, 2004: 286)

In Russell’s theory, the utterance comes out as true; Devitt claims that it is false. But Devitt’s intuition seems mistaken. If the speaker uses the quoted sentence referentially, then his utterance is successful only if his audience in some way grasps the referential proposition intended. In other words, if the utterance is referential, it cannot be understood unless the audience has cognitive contact of some sort with the referent. Yet it seems clear that the hearer can fully understand the utterance even without any such contact with the speaker’s referent. So, it would seem, the case against Russell fails. (See further Ludlow and Neale, 1991.)

An Alternative Nonreferential Account

Recently, theorists have considered a third option: that indefinites lack “quantificational force of their own” (Chierchia, 1995: 11). In this view, an indefinite description is not a quantifier, nor is it a referring expression. Rather, it resembles a free variable in that it can be bound by a quantifier that c-commands it, with the significant difference that, when not c-commanded by a quantifier, it is interpreted as bound by an existential quantifier (on c-command, see *Scope and Binding: Semantic Aspects*). For example, consider (5a):

- (5a) Whenever a man owns a donkey, he beats it.

Here, *whenever* has universal force, binding any free variables within its scope. This reading is captured in (5b):

- (5b) $\forall x, y (\text{man}(x) \ \& \ \text{donkey}(y) \ \& \ \text{owns}(x, y) \supset \text{beats}(x, y))$

A variant of this example – *if a man owns a donkey, he beats it* – receives the same analysis since, in the view in question, when an *if* clause lacks an overt quantifier, an unrealized universal quantifier is assumed to be present.

As indicated, when an indefinite is not within the scope of a quantifier, it is interpreted as bound by an existential quantifier:

- (6a) Whenever a man owns a donkey, he beats it with a stick.
 (6b) $\forall x, y ((\text{man}(x) \ \& \ \text{donkey}(y) \ \& \ \text{owns}(x, y)) \supset \exists z (\text{stick}(z) \ \& \ \text{beat-with}(x, y, z)))$

The analysis yields some counterintuitive results, however. Consider:

- (7a) If Mary finds a copy of *Middlemarch* at Bookworms, she'll buy it.
 (7b) $\forall x ((\text{copy-of-Middlemarch}(x) \ \& \ \text{finds-at}(\text{Mary}, x, \text{Bookworms})) \supset \text{buys}(\text{Mary}, x))$

In the suggested analysis, this sentence is true just in case Mary buys *every* copy of *Middlemarch* she finds at Bookworms. But, intuitively, it can be true if, finding several copies, Mary, sensibly, buys just one.

One recent response to this difficulty is to take indefinites as introducing a choice function – that is, a function from a predicate *P* to a specific member of *P*'s extension (see Reinhart, 1997). Informally, the choice-function analysis of the previous example is as follows: for some choice function *f*, Mary buys *f*(copy-of-*Middlemarch*(*x*) & finds-at(*Mary*, *x*, Bookworms)). This captures the desired truth conditions, but it seems unintuitive as an account of the literal meaning of the relevant sentence. It might, for example, be objected that speakers assertively uttering (7a) do not take themselves to be referring to, or quantifying over, functions; but this is precisely what the analysis implies.

There is, it should be added, a competing Russellian account in the literature, according to which unbound anaphora are concealed definite descriptions. This view has the potential to provide a truth-conditionally-adequate approach to the data without the awkward commitments of the current approach. (See Neale, 1990: chapters 5 and 6; also see **Anaphora: Philosophical Aspects.**)

See also: Anaphora: Philosophical Aspects; Compositionality: Philosophical Aspects; Direct Reference; Logical Form in Linguistics; Modal Logic; Pragmatic Determinants of What Is Said; Presupposition; Propositional Attitude Ascription: Philosophical Aspects; Quantifiers: Semantics; Reference: Philosophical Theories; Referential versus Attributive; Scope and Binding: Semantic Aspects.

Bibliography

- Bach K (1988). *Thought and reference*. New York: Oxford University Press.
- Chierchia G (1995). *Dynamics of meaning*. Chicago: University of Chicago Press.
- Devitt M (2004). 'The case for referential descriptions.' In Bezuidenhout A & Reimer M (eds.) *Descriptions and beyond*. Oxford: Oxford University Press. 280–305.
- Donnellan K (1966). 'Reference and definite descriptions.' *Philosophical Review* 75, 281–304. [Reprinted in Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: The MIT Press.]
- Evans G (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Graff D (2000). 'Descriptions as predicates.' *Philosophical Studies* 102, 1–42.
- Kripke S (1977). 'Speaker's reference and semantic reference.' In French P A, Uehling T E & Wettstein H (eds.) *Contemporary perspectives in the philosophy of language*. Minneapolis: University of Minnesota Press. 6–27. [Reprinted in Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: The MIT Press.]
- Lambert K (2003). *Free logic*. Cambridge: Cambridge University Press.
- Lewis D (1979). 'Scorekeeping in a language game.' *Journal of Philosophical Logic* 8, 339–359.
- Ludlow P & Neale S (1991). 'Indefinite descriptions: in defense of Russell.' *Linguistics and Philosophy* 14, 171–202.
- Makin G (2000). *The metaphysicians of meaning*. New York: Routledge.
- Mates B (1973). 'Descriptions and reference.' *Foundations of Language* 10, 409–418.
- May R (1985). *Logical form: its structure and derivation*. Cambridge: The MIT Press.
- Neale S (1990). *Descriptions*. Cambridge: The MIT Press.
- Neale S (2004). 'This, that, and the other.' In Bezuidenhout A & Reimer M (eds.) *Descriptions and beyond*. Oxford: Oxford University Press. 68–182.
- Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: The MIT Press.
- Reinhart T (1997). 'Quantifier scope: how labor is divided between QR and choice functions.' *Linguistics and Philosophy* 20, 335–397.
- Russell B (1903). *The principles of mathematics*. London: George Allen & Unwin.
- Russell B (1905). 'On denoting.' *Mind* 14, 479–493. [Reprinted in Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: The MIT Press.]
- Schiffer S (1994). 'Descriptions, indexicals, and belief reports: some dilemmas (but not the ones you expect).' *Mind* 104, 107–131. [Reprinted in Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: The MIT Press.]
- Strawson P F (1950). 'On referring.' *Mind* 54, 320–344. [Reprinted in Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: The MIT Press.]
- Szabó Z (2000). 'Descriptions and uniqueness.' *Philosophical Studies* 101, 29–57.
- Westertahl D (1986). 'Quantifiers in formal and natural languages.' In Gabbay D & Guenther F (eds.)

Handbook of philosophical logic: topics in the philosophy of language, vol. 4. Dordrecht: Kluwer. 1–131.

Wettstein H (1981). ‘Demonstrative reference and definite descriptions.’ *Philosophical Studies* 40, 241–257. [Reprinted in Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: The MIT Press.]

Whitehead A N & Russell B (1927). *Principia mathematica* (vol. I). (2nd edn.). Cambridge: Cambridge University Press. [Relevant portions reprinted in Ostertag G (ed.) (1998). *Definite descriptions: a reader*. Cambridge: The MIT Press.]

Direct Reference

A Sullivan, Memorial University of Newfoundland,
St. John's NL, Canada

© 2006 Elsevier Ltd. All rights reserved.

What Is Direct Reference?

Let us begin with a brief introduction to some essential terms. A *singular term* is an expression whose role is to specify a particular individual – proper names (e.g., ‘John’) and pronouns (e.g., ‘she’) are paradigm cases. A *proposition* is the meaning expressed by a sentence. A sentence whose subject-expression is a singular term (e.g., ‘John is tall,’ ‘She is happy’) expresses a *singular proposition*. (In contrast, a sentence whose subject-expression is a general term expresses a *general proposition* [e.g., ‘Tigers are mammals’].)

Direct reference is a contrastive term; the contrasting view is ‘mediated’ or ‘indirect’ reference. (For brevity, I’ll use the labels ‘DR’ for the former and ‘IR’ for the latter.) The classic IR position is developed by Frege (1892). On Frege’s view, terms express a sense that determines a referent. All IR views posit some such semantic mediator between words and referents, and the characteristic virtue of the IR approach is that it affords a clear explanation of how co-referential terms can differ in meaning. (For other examples of influential IR views, see Carnap [1947] and Searle [1983].)

One finds a very different approach to the semantics of (at least some) singular terms in the work of Mill and Russell. However, even though both Mill (1843: 20) and Russell (1919: 283) use the expression ‘direct’ in relevant contexts, the term ‘direct reference’ was first explicitly coined, and given a precise sense, by Kaplan (1977: 483). On Kaplan’s usage, DR expressions do not conform to one particular tenet of the IR approach, according to which the semantic mediator, the sense or the manner in which the term presents its referent, is a constituent of the proposition expressed. Kaplan’s (1977) central thesis is that indexical expressions (such as ‘I,’ ‘yesterday,’ or ‘that duck’) are DR. He also suggests (1977:

558–563) that proper names are DR. A central component of Kaplan’s picture is an account of a kind of meaning he calls ‘character’ (1977: 505–507), which is intended to explain why it seems that co-referential DR terms can make distinct contributions to propositional content.

The heart of Kaplan’s notion of DR is Russell’s criterion for individuating singular propositions: “A name is merely a means of pointing to the thing, and does not occur in what you are asserting, so that if one thing has two names you make exactly the same assertion, whichever of the names you use...” (Russell, 1918: 245). If a term is DR, in Kaplan’s sense, then the propositions expressed by sentences in which it figures are individuated solely in terms of the individuals and properties that they are about, as opposed to in terms of more finely grained senses or concepts. Because the contribution a DR term makes to propositional content is its referent, sentences that differ only in the interchange of co-referential DR terms express the same proposition.

The difference between the DR and IR views, then, is most stark concerning such pairs of sentences – e.g., ‘That [pointing to the heavens] is Mars’ versus ‘Mars is Mars.’ On the IR view, there is a clear difference between the propositions expressed by such pairs of sentences; for one has as a constituent the meaning or sense of ‘that,’ whereas the other has as a constituent the meaning or sense of ‘Mars.’ However, the price paid is that IR views posit mediators between terms and referents, and critics allege that these mediators create more problems than they solve. On the DR view, there are no semantic mediators – no senses or concepts – involved in the content of a singular proposition; proponents of the view argue that this affords a more satisfactory account of the content and truth-conditions of such propositions. (That is, singular propositions are about the relevant referents *per se*, not about whatever might happen to satisfy a certain sense or concept.) However, the DR view allows no room for the intuition that such pairs of sentences can express distinct propositions: if ‘that’ and ‘Mars’ are co-referential, there is

no semantic difference between the propositions expressed by sentences which differ only in their interchange. (See later for more on this difference, and its consequences.)

Recanati (1993) is an important subsequent work on DR. (Recanati is heavily influenced by the works of Perry, collected in Perry [1993].) In the interim, Evans (1982) and McDowell (1986) had spurred a neo-Fregean approach to singular terms by arguing that many criticisms of Frege's views can be met if senses are conceived as object-dependent (and so 'rigid,' in Kripke's [1972] terminology). Recanati defines DR terms as those with a semantic feature that indicates that the truth-condition expressed by sentences containing them is singular or object-dependent, and argues that this gets to the core of the difference between referring expressions (such as names and indexicals) and quantified noun phrases (such as 'a man' or 'all Texans'). Recanati's notion of DR is weaker than Kaplan's, in that neo-Fregean singular terms (which express object-dependent senses) would be classified as DR in Recanati's sense but not in Kaplan's (because object-dependent senses figure as constituents of propositions for Recanati, which is inconsistent with Kaplan's Russellian criterion for individuating propositions). On Recanati's view, as distinct from Kaplan's, sentences that differ only in the interchange of co-referential DR terms, although truth-conditionally equivalent, express distinct propositions.

Neither Kaplan nor Recanati deny that DR expressions are semantically associated with something like a sense or manner of presentation. (This can hardly be denied for indexicals.) What Kaplan explicitly denies, in calling an expression DR, is that the sense or manner of presentation affects propositional content. What Recanati explicitly denies is that the sense or manner of presentation is truth-conditionally relevant. So, strictly speaking, the contemporary authorities use 'direct reference' to label an approach to propositional content, more so than as a label for any specific approach to reference.

Some Closely Related Concepts

Kaplan's DR is a part of an anti-Fregean tide that swept the philosophy of language in the 1970s. In much of the secondary literature, the distinction between DR and other aspects of that movement – such as the causal-historical theory of reference, the notion of rigid designation, and the Millian view of proper names – is lamentably blurred. It is not uncommon to find a bundle of such notions lumped together under the umbrella term 'the new theory of reference.' The aim of this section is to be more discriminating about

the relations between DR and these other concepts and views.

The Millian view has it that names are connotationless labels: "A proper name is but an unmeaning mark which we connect in our minds with the idea of the object..." (Mill, 1843: 22). This is no mere claim about truth-conditions; it is a very strong claim about the semantics of names. Mill even says that: "The name, therefore, is said to signify the subjects *directly*..." (1843: 20). Nonetheless, it is clear that Mill's notion of direct reference is quite distinct from Kaplan's. Witness the fact that indexical expressions (e.g., 'I,' 'yesterday') are the very paradigm of DR for Kaplan, but are clearly not unmeaning marks. Kaplan (1977: 520) discusses this "drawback to the terminology 'direct reference'" – that it "suggests falsely that the reference [of an indexical] is not mediated by the meaning, which it is." Explicitly, Kaplan (1989: 568) does not deny that DR terms are semantically associated with something like a sense. What he denies is that the sense is a propositional constituent.

In the case of proper names, though, this conflation of DR with the Millian view is more prevalent, and more difficult to rebut. One reason is that there are strong considerations against the view that names are semantically associated with a particular semantic mediator (see especially Donnellan [1970] and Kripke [1972]). So, in the case of names, as compared with indexicals, it is more difficult to identify anything that plays a semantic role akin to Kaplan's character, singling out the referent without affecting propositional content. Hence, although it is implausible to hold that indexicals are Millian (even if they are DR), it is not uncommon to encounter the idea that names are Millian (i.e., a name's meaning is just its referent) and that names are DR (i.e., a name's contribution to propositional content is just its referent) are two sides of the one coin. However, to the contrary, there is clearly conceptual space for the view that proper names are DR but not Millian (i.e., names are semantically associated with some kind of sense or meaning, but nonetheless that sense or meaning is truth-conditionally irrelevant, no part of propositional content).

Although a term could clearly be DR without being Millian, it is plausible that if a term is Millian then it is DR. That is, if all that there is to the semantics of a term is that it labels a specific referent, then it is hard to see what else but the referent could affect propositional content. A similar relation holds between Kaplan's DR and Kripke's (1972) rigid designation. (A designator is rigid if it designates the same thing in every possible world.) There are clearly rigid designators that are not DR (say, 'the even prime'), but nonetheless any DR term will satisfy the criterion

for rigidity (see Kaplan [1977: 492–498, 1989: 569–571]). So, there are rigid terms that are not DR, and DR terms that are not Millian; but any Millian term would be DR, and all DR terms are rigid.

There is a nice fit between the semantics of DR and the causal-historical story of how reference is determined. Donnellan's (1970) and Kripke's (1972) influential arguments that the meaning of a proper name is not some kind of descriptive sense have been taken to demonstrate, or to suggest, general and fundamental problems with any IR approach to names. From here, it looks compelling to conclude that a name's contribution to propositional content is just its referent. Furthermore, it is possible that appeal to distinct causal-historical chains of transmission can explain why it seems that co-referential names differ in meaning. (More on this later.)

In any case, the present point is just that these are all are distinct doctrines, addressed to quite different questions. The Millian view is a bold conjecture about the semantics of names, DR is a somewhat weaker claim about propositional content, rigid designation is a still weaker modal claim about certain terms, and the causal-historical theory is a picture of how reference is determined. There are deep and interesting relations between these concepts and views, but they should not be conflated.

Problems with Direct Reference

The classic problems for DR are, not surprisingly, the very problems that led to the development of the IR view in the first place. The central problem concerns questions of substitutivity – that is, there are reasons to think that interchanging co-referential expressions fails to preserve propositional content. Substitutivity arguments against the DR view of names have been part of the canon since Frege (1892). Competent speakers can fail to recognize that co-referential names do in fact name the same object, and so sentences that differ only in the interchange of co-referential names seem to express distinct propositions. Clearly, similar things can happen with indexicals. Consider, for example, Kaplan's (1977: 537) case wherein he sees a reflected image of a man whose pants are on fire, then subsequently come to recognize that he is that man. Even though his initial 'His pants are on fire' and subsequent 'My pants are on fire' are truth-conditionally equivalent, there are significant semantic differences between them.

So, the DR theorists need to accommodate the considerable reasons for thinking that sentences that differ only in the interchange of co-referential terms can express distinct propositions. Alternatively put, although the IR view issues in more finely grained

propositions, which are better suited to capture the content of a belief, the DR view issues in propositions which are too coarsely grained for this purpose. (For example, believing that 'His pants are on fire' and believing that 'My pants are on fire' would prompt different actions; insofar as the DR view is committed to the claim that these beliefs have the same content, then, something is amiss.)

Kaplan's (1977) notion of 'character' is intended to solve this problem for the case of indexicals – i.e., to explain the differences between 'His pants are on fire' and 'My pants are on fire,' in a way that is consistent with the semantics of DR. The extent to which Kaplan's account is successful is one of the central points of controversy in the subsequent literature on this topic. Note also that there are cases, first raised by Perry (1977), which character cannot accommodate. The cases concern two uses of the same demonstrative – and so character remains constant – wherein, unbeknownst to the speaker, the same individual is referred to twice. (For instance, a speaker utters 'This ship (pointing to the stern of a ship through one window) is American but this ship (pointing to the bow of a ship through a different window) is Japanese.') Kaplan (1977: 514ff) makes some effort toward accommodating this kind of problem, but his remarks are sketchy, and have been contested. (Cf. Braun [1994] for discussion, and King [2001] for an overview of the burgeoning literature.)

In the case of names, DR theorists have drawn on the causal-historical theory of reference to explain why it seems that co-referential names differ in meaning. This explanation, developed by Kripke (1972), Kaplan (1989), and Stalnaker (1997), relies on a sharp distinction between the semantic question of what a word refers to and the metasemantic question of why a word refers to what it does. DR is a semantic claim (i.e., sentences containing names express Russellian singular propositions), and the causal-historical theory suggests a complimentary metasemantic claim (i.e., co-referential names are distinct words with different causal histories, and so there may be all manner of non-semantic differences between them). The thought is that these (non-semantic) differences can explain why uses of co-referential names can communicate different things, even though the names are semantically equivalent.

Whether such an account holds any promise of saving DR from Frege's problem is also much contested. For attempts to make the case, see Salmon (1986) and Soames (2002); for arguments against its promise, see Schiffer (1987, 2003: Chap. 2). These and other debates surrounding DR continue to be among the most vibrant in contemporary philosophy of language.

See also: Causal Theories of Reference and Meaning; Character versus Content; Empty Names; Object-Dependent Thoughts; Proper Names: Philosophical Aspects; Reference: Philosophical Theories; Rigid Designation; Sense and Reference: Philosophical Aspects.

Bibliography

- Braun D (1994). 'Structured characters and complex demonstratives.' *Philosophical Studies* 74, 193–219.
- Carnap R (1947). *Meaning and necessity*. Chicago: University of Chicago Press.
- Donnellan K (1970). 'Proper names and identifying descriptions.' *Synthese* 21, 256–280.
- Evans G (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Frege G (1892). 'On sense and reference.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 175–192.
- Kaplan D (1977). 'Demonstratives.' In Almog J, Perry J & Wettstein H (eds.) *Themes from Kaplan*. Oxford: Oxford University Press, 1989. 481–564.
- Kaplan D (1989). 'Afterthoughts.' In Almog J, Perry J & Wettstein H (eds.) *Themes from Kaplan*. Oxford: Oxford University Press. 565–614.
- King J (2001). *Complex demonstratives*. Cambridge, MA: MIT Press.
- Kripke S (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.

- MacDowell J (1986). 'Singular thought and the extent of inner space.' In MacDowell J & Pettit P (eds.) *Subject, thought, and context*. Oxford: Oxford University Press. 137–168.
- Mill J S (1843). *A system of logic*. London: Longmans.
- Perry J (1977). 'Frege on demonstratives.' *Philosophical Review* 86, 474–497.
- Perry J (1993). *'The Problem of the Essential Indexical' and other essays*. Oxford: Oxford University Press.
- Recanati F (1993). *Direct reference*. Oxford: Blackwell.
- Russell B (1918). The philosophy of logical atomism. In Marsh R C (ed.) *Logic and Knowledge*. London: Unwin Hyman, 1956. 175–282.
- Russell B (1919). 'Descriptions.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 279–287.
- Salmon N (1986). *Frege's puzzle*. Cambridge, MA: MIT Press.
- Schiffer S (1987). 'The 'Fido'-Fido theory of belief.' *Philosophical Perspectives* 1, 455–480.
- Schiffer S (2003). *The things we mean*. Oxford: Oxford University Press.
- Searle J (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Soames S (2002). *Beyond rigidity*. Oxford: Oxford University Press.
- Stalnaker R (1997). 'Reference and necessity.' In Hale B & Wright C (eds.) *A companion to the philosophy of language*. Oxford: Blackwell. 534–553.

Discourse Representation Theory

J van Eijck, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands and Research Institute for Language and Speech, Utrecht, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

The Problem of Unbound Anaphora

The most straightforward way to establish links between anaphoric pronouns and their antecedents is to translate the pronouns as variables bound by their antecedents. This approach does not work when the link crosses a sentence boundary, as in example (1).

- (1) *A man₁ met an attractive woman₂. He₁ smiled at her₂.*

It should be possible to interpret the first sentence of this discourse as soon as it is uttered, and then later on, while processing the second sentence, establish the links between the pronouns and their intended antecedents.

One possible solution is translating the indefinites by means of existential quantifiers with scopes

extending beyond the sentence level and then allowing the variables for the pronouns to be captured by these quantifiers. But this will not do: at some point the scope of a quantifier must be 'closed off,' but further on another pronoun may occur that must be linked to the same antecedent.

The bound variable approach to anaphora also fails for cases where a pronoun in the consequent of a conditional sentence is linked to an indefinite noun phrase in the antecedent of the conditional, as in example (2).

- (2) *If a man₁ meets an attractive woman₂, he₁ smiles at her₂.*

A possible approach here would be to view (2) as a combination of the noun phrases *a man* and *an attractive woman* with a structure containing the appropriate gaps for antecedents and pronouns, viz., (3). This is the approach of *quantifying-in*, taken in traditional Montague grammar (see **Montague Semantics**).

- (3) *If PRO₁ man meets PRO₂, PRO₁ smiles at PRO₂.*

This approach does not work here, however. Quantifying-in the indefinite noun phrases in (3), i.e., in a structure that has the conditional already in place, would assign the wrong scope to the indefinites with respect to the conditional operator.

Note that the meaning of (2) is approximately the same as that of (4).

- (4) *Every man who meets an attractive woman₁ smiles at her₁.*

In this case as well, quantifying-in does not allow one to generate the most likely reading where the subject of the sentence has wide scope over the embedded indefinite. Sentences with the patterns of (2) and (4) have reached the modern semantic literature through Geach (1962). Geach's discussion revolves around examples with donkeys, so these sentences became known in the literature as '*donkey sentences*.'

As has repeatedly been remarked in the literature, there are quite striking structural parallels between nominal and temporal anaphora. The past tense can be viewed as an anaphoric element in all those cases where it is not to be understood as 'sometime in the past' but as referring to some definite past time.

- (5) *John saw Mary. She crossed the street.*

In example (5), presumably the seeing takes place at some specific time in the past, and the crossing takes place immediately after the seeing. Again, we have an anaphoric link across sentence boundaries, and a traditional operator approach to tense does not seem to fit the case. Although tense is not treated in the pioneer papers on discourse representation, it is clear that the problem of temporal anaphora is a very important subproblem of the general anaphora problem that discourse representation theory sets about to solve.

Basic Ideas

Discourse representation theory as it was presented in Kamp (1981) addressed itself specifically at the problem of the previous section, although confined to nominal anaphora. The basic idea of the approach is that a natural language discourse (a sequence of sentences uttered by the same speaker) is interpreted in the context of a representation structure. The result of the processing of a piece of discourse in the context of representation structure R is a new representation structure R' ; the new structure R' can be viewed as an updated version of R .

The interpretation of indefinite noun phrases involves the introduction of '*discourse referents*' or '*reference markers*' for the entities that a piece of discourse is about. In the following, the term

'*discourse referent*' will be used. Discourse referents are essentially free variables. Thus, indefinite noun phrases are represented without using existential quantifiers. The quantification is taken care of by the larger context. It depends on this larger context whether an indefinite noun phrase gets an existential reading or not.

The life span of a discourse referent depends on the way in which it was introduced. All 'alive' referents may serve as antecedents for anaphors in subsequent discourse. Anaphoric pronouns are represented as free variables linked to appropriate antecedent variables. Definite descriptions in their simplest use are treated in a way that is similar to the treatment of anaphoric pronouns: definite noun phrases in their anaphoric use are treated like indefinite noun phrases; i.e., they are translated as free variables, but give rise to additional anaphoric links. The treatment of other, functional uses of definite noun phrases (as in *A car crashed. The driver emerged unharmed.*) is more involved.

The difference between indefinite noun phrases, on the one hand, and definite noun phrases and pronouns, on the other, is that indefinites introduce *new* variables, whereas the variables introduced by definites and pronouns are always linked to an already established context. In other words, the difference between definites (including pronouns) and indefinites is that the former refer to entities that have been introduced before, i.e., to *familiar* entities, whereas the latter do not.

Quantifier determiners, i.e., determiners of noun phrases that are neither definite nor indefinite, can bind more than one variable. Specifically, they can bind a *block* of free variables, some of which may have been introduced by indefinites. Conditional operators (*if ... then ...* constructions) can also bind blocks of free variables. Not all variables introduced by indefinites are in the scope of a quantifier or a conditional operator. Those that are not are existentially quantified over by default.

The processing of a piece of discourse is incremental. Each next sentence to be processed is dealt with in the context of a structure that results from processing the previous sentences. The processing rules decompose a sentence, replacing the various parts by *conditions* to be added to the structure. Assume one is processing discourse (6) in the context of representation structure (7) containing just one discourse referent and one condition.

- (6) *A man walked down the street. He whistled.*

- (7) (x) (*street* (x))

As was mentioned before, indefinite noun phrases give rise to new discourse referents, and definite noun

phrases are linked to existing discourse referents. The indefinite in the first sentence of (6) introduces a new discourse referent y and two conditions $man(y)$ and y *walked down the street*. The second condition can be decomposed further by introducing a fresh discourse referent in the structure, linking this discourse referent to an existing discourse referent, and replacing the definite noun phrase with the discourse referent in two new conditions. This gives three new conditions all in all: $z = x$, *street*(z) and *walked-down*(y , z). The discourse representation structure now looks like

- (8) (x, y, z) (*street*(x), *man*(y), $z = x$, *street*(z),
walked-down(y , z))

Processing the second sentence of (1) gives rise to a new link and a new condition. The final result is (9).

- (9) (x, y, z, u) (*street*(x), *man*(y), $z = x$, *street*(z),
walked-down(y , z), $u = y$, *whistled*(u))

All representation conditions in the above example are atomic. Quantified noun phrases or logical operators, such as conditionals or negations, give rise to complex conditions. The representation structure for (4) given in (10) provides an example.

- (10) $((x, y)$ (*man*(x), *woman*(y), *attractive*(y),
meet(x , y))) \Rightarrow $((), (\textit{smiles-at}$ (x , y)))

Note the appearance of an arrow \Rightarrow between components of the structure, glueing two nonatomic pieces of representation together. Note also that the right-hand component starts with an empty list $()$, to indicate that on the right-hand side no new discourse referents are introduced.

In the box format that many people are perhaps more familiar with, (10) looks like (11).

- (11)

x	y
<i>man</i> x	
<i>woman</i> y	
<i>attractive</i> y	
<i>meet</i> (x , y)	

 \Rightarrow

<i>smiles-at</i> (x , y)

Formal definitions and truth conditions for these representation structures are given in the next section.

Kamp (1981) and Kamp and Reyle (1990) spell out the rules for processing sentences in the context of a representation structure in all the required formal detail. An important feature of the rules is that they impose formal constraints on availability of discourse referents for anaphoric linking. Roughly, the set of available discourse referents consists of the discourse referents of the current structure, plus the discourse referents of structures that can be reached from the current one by a series of steps in the directions *left* (i.e., from the consequent of a pair $R \Rightarrow R'$ to the antecedent), and *up* (i.e., from a structure to an

encompassing structure). The constraints on discourse referent accessibility are used to explain the awkwardness of anaphoric links, as in (12).

- (12) *If every man_1 meets an attractive $woman_2$, he_1 smiles at her_2 .

Such data can be disputed, but space does not permit such indulgence here. Discourse referents for proper names are always available for anaphoric reference; to reflect this fact, such discourse referents are always included in the list of discourse referents of the top-level structure.

To account for deictic uses of pronouns, use is made of *anchored structures*. An anchored structure is a pair consisting of a representation structure R and a function f , where f is an *anchor* for a subset of the discourse referents of R ; i.e., f assigns appropriate individuals in a model to these discourse referents. For example, structure (7) could be anchored by mapping discourse referent x to an appropriate street. Deictic pronouns are handled by linking them to anchored discourse referents.

Essentially the same approach to natural language analysis as was proposed in Kamp (1981) is advocated in Heim (1982). Heim uses the metaphor of a filing cabinet: the established representation structure R is a file, and additions to the discourse effect a new structure R' , which is the result of changing the file in the light of the new information (see **Dynamic Semantics**).

The main program of discourse representation theory (in its generic sense) is an attempt to regard semantic interpretation as a dynamic process mapping representations plus contexts to new representations plus contexts. As Partee (1984) remarked, this shift from static semantics to dynamic semantics *cum* pragmatics means an enrichment of the enterprise of formal semantics and should therefore make it easier to establish contact with other schools of semantics and/or pragmatics. Partee's prediction was proved correct in subsequent years by the widespread use of discourse representation theory in computational linguistics and by the application of techniques of anaphora resolution from Artificial Intelligence in systems based on discourse representation theory.

Discourse representation theory has also provided new inspiration to traditional Montague grammarians, who tend to be less than satisfied with the contextual rules for analyzing discourse on the grounds that the influence of context makes it difficult to work out what contribution individual phrases make to the meaning of the whole. A suitable dynamic perspective on the process of interpretation has shown these compositionality qualms to be unfounded, and discourse representation theory has been instrumental

in bringing about this dynamic turn (see **Dynamic Semantics** for details).

Heim (1990) contains a perceptive appraisal of various alternatives to the approach of discourse representation theory (in its generic sense) to the problem of unbound anaphora.

Discourse Representation Structures (DRSs)

Formally, a discourse representation structure R consists of two parts: a finite list of *discourse referents* and a finite list of *conditions*. The discourse referents in the list are called the discourse referents of R . The conditions of a structure R may contain discourse referents that are not included in the list of discourse referents of R . Conditions can be *atoms*, *links*, or *complex conditions*. An atom is a predicate name applied to a number of discourse referents, a link is an expression $t = r$, where r is a discourse referent and t is either a proper name or a discourse referent. The clause for complex conditions uses recursion; a complex condition is a condition of the form $R \Rightarrow R'$, where R and R' are discourse representation structures.

Next, one defines truth for discourse representation structures with respect to a model. Call $M = \langle D, I \rangle$ an appropriate model for discourse representation structure R if I maps the discourse referents of R to members of D , the n -place predicate names in the atomic conditions of R to n -place relations on D , the names occurring in the link conditions of R to members of D , and (here is the recursive part of the definition) M is also appropriate for the structures in the complex conditions of R .

Let $M = \langle D, I \rangle$ be an appropriate model for structure R . An assignment in $M = \langle D, I \rangle$ is a mapping of discourse referents to elements of D . Assignment f verifies R in M if there is an extension f' of f with the following properties:

1. f' is defined for all discourse referents of R and for all discourse referents occurring in atomic or link conditions of R .
2. If $P(r_1, \dots, r_n)$ is an atomic condition of R , then $\langle f'(r_1), \dots, f'(r_n) \rangle \in I(P)$.
3. If $t = r$ is a link condition of R , and t and r are both discourse referents, then $f'(t) = f'(r)$; if t is a proper name and r a discourse referent, then $I(t) = f'(r)$.
4. If $R_1 \Rightarrow R_2$ is a complex condition of R , then every assignment for R_1 that verifies R_1 and agrees with f' on all discourse referents that are not discourse referents of R_1 also verifies R_2 .

A structure R is true in M if the empty assignment verifies R in M . These definitions can be modified to take anchors into account in the obvious way, by focusing on assignments extending a given anchor.

Clearly, the expressive power of this basic representation language is quite limited. In fact, there is an easy recipe for translating representation structures to formulae of first-order predicate logic. Assuming that discourse referents coincide with predicate logical variables, the atomic and link conditions of a representation structure are atomic formulae of predicate logic. The translation function $^\circ$, which maps representation structures to formulae of predicate logic, is defined as $R^\circ = \bigwedge C_i^\circ$, where \bigwedge indicates a finite conjunction and the C_i° are the translations of the conditions of R . The translation for conditions is in turn given by the following clauses.

- For atomic conditions: $C^\circ = C$.
- For complex conditions: $(R_1 \Rightarrow R_2)^\circ = \forall x_1 \dots \forall x_n (R_1^\circ \Rightarrow \exists y_1 \dots \exists y_m R_2^\circ)$, where x_1, \dots, x_n is the list of discourse referents of R_1 and y_1, \dots, y_m the list of discourse referents of R_2 .

It is easy to show that R is true in M under the definition given above if and only if R° is true in M for some assignment, under Tarski's definition of truth for first-order predicate logic.

A slight extension of the discourse representation language allows for the treatment of negation. Negated conditions take the form $\neg R$, where R is a discourse representation structure. Negations of atomic conditions are treated as negations of discourse representation structures containing just one atomic condition. The discourse referents of a negated structure are not available for anaphoric linking outside that structure.

The definition of satisfaction must take negated conditions into account. Here is the required extension of the definition. Assignment f verifies R in M if there is an extension f' of f with the following properties:

- 1–4. As above.
5. If $\neg R'$ is a complex condition of R , then no assignment that agrees with f' on all discourse referents that are not discourse referents of R' verifies R' .

Translation into predicate logic now must take care of negation as well. The translation clause for negated conditions runs as follows:

- $(\neg R)^\circ = \neg \exists x_1 \dots \exists x_n R^\circ$

Here x_1, \dots, x_n is the list of discourse referents of R . It is easy to see that the given translation is meaning-preserving. It is also not difficult to give a meaning-preserving translation in the other direction. This shows that the discourse representation language extended with negation has precisely the same expressive power as first-order predicate logic.

Extensions: Tense and Plurals

Partee (1984) gave a survey of proposals to extend discourse representation theory with discourse referents for times and events to exploit the parallels between nominal and temporal anaphora. In example (5) from section 1, where first reference is made to a seeing event in the past and then to an event of crossing the street that takes place immediately after the seeing event, an anchoring mechanism can be used to link the seeing event to the appropriate time, and an anaphoric link between events can constrain the time of the crossing event in the appropriate way. Also, the dynamic effect of shifting the reference time can be incorporated by using a designated discourse referent for the reference time and specifying that this discourse referent be updated as a side effect of the processing of sentences denoting events.

Next, there are examples where a reference is picked up to an indefinite time in the past.

- (13) *Mary arrived during the day. She let herself into the house.*

In example (13), the arrival takes place at some indefinite time on a specific day (presumably anchored) in the past. The event of Mary's entering the house is then linked to the time of arrival. Again, all that is needed is the introduction of an event discourse referent for the arrival event and an appropriate linking of this event discourse referent to the reference time discourse referent: the reference time discourse referent starts pointing at a time interval just after the time of arrival. The processing of the next sentence introduces an event that is constrained to be included in the reference time interval and has again as a side effect that the reference time discourse referent is shifted to refer to a time interval just after the house-entering event.

Sentence (14) provides an example of quantification over times.

- (14) *When Bill called, Mary was always out.*

The example gives rise to a complex representation of the form $R \Rightarrow R'$, with an event discourse referent and a reference time discourse referent introduced in the left-hand structure, and a state discourse referent of the right-hand structure, with the state constrained to include the reference time interval.

An operator account of tenses and temporal adverbs has the awkwardness that the tense operator is redundant if a temporal adverb is present, as in (15), but not otherwise. Also, assigning the correct scopes to these operators poses problems.

- (15) *Bill called last Friday around noon.*

In the discourse representation approach, where tenses translate into event or state variables linked to an appropriate reference time, temporal operators are simply translated as predications on the event discourse referent, and the awkwardness vanishes. See Kamp and Rohrer (1983) and Partee (1984), plus the references cited therein, for details.

As for the incorporation of the singular versus plural distinction, an obvious first move in any attempt to accommodate plural anaphoric pronouns is to make a distinction between singular and plural discourse referents. Singular pronouns are linked to singular discourse referents, and plural pronouns are linked to plural discourse referents. Plural indefinite noun phrases (*some women, three men*) introduce plural discourse referents, but it turns out that many other introduction mechanisms must be postulated to obtain a reasonable coverage of plural anaphoric possibilities.

Plural discourse referents may result from *summation* of singular discourse referents. This is to account for uses of *they* that pick up a reference to a set of individuals that have been introduced one by one. Next, plural individuals may be the result of *abstraction* from complex conditions. Consider example (16).

- (16) *John bought every book Mary had mentioned. He started reading them straight away.*

Obviously, *them* refers to the set of all books mentioned by Mary. No plural discourse referent is introduced by the first sentence, so the only way to make one available is by calling it into being through abstraction.

So-called *dependent plurals* should be handled differently again, because here the plurality seems closely linked to syntax. Sentence (17) provides an example.

- (17) *All my friends have children.*

It is clear that (17) is still true if each of my friends has exactly one child. Dependent plurals call for a kind of in-between discourse referent that is neutral between singular and plural. The chapter on plurals in Kamp and Reyle (1990) gives a very detailed account of these and related matters. **Plurality** provides further information on general issues of the interpretation of plurals.

Incorporating Generalized Quantifiers

Extending discourse representation theory with nonstandard quantifiers, and then getting the truth conditions right, is not completely straightforward.

- (18) *Most farmers who own a donkey beat it.*

Applying a routine strategy for building a representation structure for example (18), one arrives at

structure (19), where $R \Rightarrow_m \Rightarrow R'$ is true if most verifying assignments for R are verifying assignments for R' .

$$(19) ((x, y)(\text{farmer}(x), \text{donkey}(y), \text{own}(x, y))) \Rightarrow_m \Rightarrow ((\))(\text{beat}(x, y)))$$

This analysis does give the wrong truth conditions, because it quantifies over farmer–donkey pairs instead of individual farmers. In a situation where there are five kind farmers who each own 1 donkey and treat it well, and one cruel, rich farmer who beats each of his 10 donkeys, the analysis makes sentence (18) true, though intuitively it should be false in this situation.

The remedy (proposed in Kamp and Reyle, 1990) involves a complication in the notation. Generalized quantifiers are introduced explicitly in the representation structures. The revised representation for (18) is (20).

$$(20) ((x, y)(\text{farmer}(x), \text{donkey}(y), \text{own}(x, y))) \Rightarrow_x^{\text{most}} \Rightarrow ((\))(\text{beat}(x, y)))$$

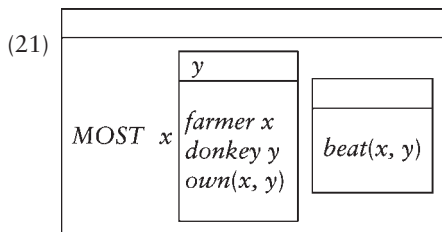
At the place of *most* in (20) one could in principle have any generalized quantifier (see **Quantifiers: Semantics**). In other words, for every binary generalized quantifier Q and every pair of representation structures R, R' , the following is a complex condition: $R \Rightarrow_v^Q \Rightarrow R'$. The truth conditions are modified to reflect what is expressed by the quantifier Q . Generalized quantifiers express relations between sets, so $R \Rightarrow_x^Q \Rightarrow R'$ is true in case the two sets are in the appropriate quantifier relation. The truth conditions must pick out the two relevant sets. Here is the new part of the definition. Assignment f verifies R in M if there is an extension f' of f with the following properties:

1–5. As above.

6. If $R_1 \Rightarrow_v^Q \Rightarrow R_2$ is a complex condition of R , then f' verifies this condition if the sets B and C are in the quantifier relation denoted by Q , where $B = \{b \mid f' \text{ has an extension } g \text{ with } g(v) = b, \text{ which verifies } R_1 \text{ in } M\}$ and $C = \{c \mid f' \text{ has an extension } h \text{ with } h(v) = c, \text{ which verifies } R_2 \text{ in } M\}$.

It is left to the reader to check that this gets the truth conditions for (18) correct.

The following representation of (18) brings the incorporation of generalized quantifiers more in line with standard logical notation:



Discourse Structures and Partial Models

There is more than an occasional hint in the original papers of Kamp and Heim that discourse representation structures are closely connected to *partial* models. If the suggestion is not that these representation structures are themselves partial models, it is at least that they are intended to be interpreted with respect to partial models. That the structures are themselves partial models cannot be right: complex conditions are *constraints* on models rather than model components. They specify general properties that a model must satisfy the condition. Interpretation of discourse representation structures in partial models has never really been worked out. The truth definitions for representation structures, e.g., in Heim (1982), Kamp (1981), Kamp and Reyle (1990), define satisfaction in classical (i.e., ‘complete’) models.

Because the representation structures contain identity links and negated identity links, evaluation in partial models where not only the predicates used to translate the vocabulary of the fragment, but also the identity predicate receives a partial interpretation, is feasible. Interestingly, this sheds light on some puzzling aspects of identity statements. Current studies of partial model theory interpret identity as a total predicate (see Langholm, 1988). Partializing identity leads to a more radical form of partiality; it has the effect that the objects in the model are not proper individuals but rather proto-individuals that can still fuse into the same individual after some more information acquisition about the identity relation. Technically, this form of radical partiality can be implemented by evaluating discourse representation structures with respect to models where the identity relation is a partial relation.

The formal development of a theory of partial identity involves an interpretation of identity as a pair $\langle I^+, I^- \rangle$, with I^+ an equivalence relation that denotes the positive extension of identity, and I^- an anti-equivalence relation, that is to say, a relation that is irreflexive, symmetric, and anti-transitive, i.e., satisfying the requirement that if I^-xy , then it holds for all z that I^-xz or I^-zy .

The assumption that proto-individuals rather than regular individuals populate the partial models is attractive from the point of view of knowledge representation: often human beings have only partial information about identities. Famous paradoxes and puzzles are based on this fact. One example is Frege’s morning star, evening star paradox; see the article **Coreference: Identity and Similarity**. Another is Saul Kripke’s Pierre puzzle. Pierre is a Frenchman who has read about a famous and wonderful city he knows as *Londres*, and because of his reading he thinks that

Londres is pretty. Later on, he is abducted and forced to work in a slum in a city that, as he learns, is called *London*, and this new experience leads him to conclude that *London* is ugly. The important point to note is that as long as all this information is processed with respect to a partial model where *London* and *Londres* name different proto-individuals, Pierre's beliefs are not incoherent. They only become incoherent once it is discovered that *London* and *Londres* are identical, i.e., once Pierre acquires additional information about the extension of the identity relation. From outside, from a situation where *London* and *Londres* are anchored to the same individual, the belief may seem incoherent as well, but the point is that Pierre does not have full information about the nature of this anchor. The example is discussed in the context of discourse representation theory in Asher (1986), but the solution proposed there is still phrased in terms of classical models.

Reasoning with DRSs

The plausibility of using Discourse Representation Structures to model belief and other propositional attitudes is closely connected with the existence of cognitively plausible inference systems for DRSs. Proof theories for DRSs are given in Saurer (1993), Kamp and Reyle (1996), and Van Eijck (1999). The calculus of Van Eijck (1999) is perhaps the simplest of these, and we present it here.

We switch to the version of DRT where DRS negation is primitive and $D_1 \Rightarrow D_2$ is defined in terms of negation. The precise definition is given below. A slight modification of the DRS definition is to make a distinction between the *fixed* discourse referents and the *introduced* discourse referents of a DRS (first proposed in Visser 1994). This allows for a natural definition of DRT consequence. If a DRS is inferred, its fixed discourse referents are supposed to be supplied by the premises of the inference. They are 'fixed by the context of the inference,' so to speak.

Thus, we view a DRS as a triplet consisting of a set of fixed referents F , a set of introduced referents I , and a set of conditions $C_1 \dots C_n$ constrained by the requirement that the free variables in C_i must be among $F \cup I$. Concretely, the syntax of DRT looks like this (equality statements left out for simplicity of exposition):

Definition 1 (DRT)

$$\begin{aligned} t &::= c \mid v \\ C &::= T \mid Pt_1 \dots t_n \mid \neg D \\ D &::= \frac{v_1 \dots v_n \quad v_{n+1} \dots v_m}{C_1 \dots C_k} \end{aligned}$$

We will use \perp as an abbreviation of $\neg \top$. The (active) discourse referents of a term t or condition C or DRS D are given by:

$$\begin{aligned} M(v) &:= \{v\}, M(c) := \emptyset, M(T) := \emptyset, M(Pt_1 \dots t_n) := \cup_i M(t_i), \\ M\left(\frac{F \quad I}{C}\right) &:= F \cup I, M\left(\neg \frac{F \quad I}{C}\right) := F \end{aligned}$$

Conditions on the formation rule for a DRS

$$\frac{v_1 \dots v_n \quad v_{n+1} \dots v_m}{C_1 \dots C_k}$$

1. $\{v_1 \dots v_n\} \cap \{v_{n+1} \dots v_m\} = \emptyset$.
2. $\cup_i M(C_i) \subseteq \{v_1 \dots v_m\}$

We define the condition

$$\frac{F \quad I}{C_1 \dots C_n} \Rightarrow D$$

as

$$\neg \frac{F \quad I}{C_1 \dots C_n \neg D}$$

Here is a semantics for DRT in terms of partial assignments, following the original set-up in Kamp (1981).

Definition 2 (Semantics of DRT)

$$\begin{aligned} \mathcal{M}, f &\models T && \text{always} \\ \mathcal{M}, f &\models Pt_1 \dots t_n && \text{iff } \langle [t_1]_f^{\mathcal{M}}, \dots, [t_n]_f^{\mathcal{M}} \rangle \in P^{\mathcal{M}} \\ \mathcal{M}, f &\models \neg D && \text{iff there is no } g \text{ with } \mathcal{M}, f, g \models D \end{aligned}$$

$$\mathcal{M}, f, g \models \frac{F \quad I}{C_1 \dots C_n} \text{ iff } \begin{aligned} &f : F \rightarrow \text{dom}(\mathcal{M}), \\ &g : F \cup I \rightarrow \text{dom}(\mathcal{M}), f = g \upharpoonright F, \\ &M, g \models C_1, \dots, M, g \models C_n. \end{aligned}$$

Here $g \upharpoonright F$ denotes the restriction of function g to the set F . The following definition of DRT consequence makes essential use of the distinction between fixed and introduced discourse referents.

Definition 3 (DRT Consequence)

$$\frac{F \quad I}{C_1 \dots C_n} \models \frac{F \cup I \quad I}{C_{n+1} \dots C_m}$$

$$\text{iff for all } \mathcal{M}, f, g \text{ with } \mathcal{M}, f, g \models \frac{F \quad I}{C_1 \dots C_n}$$

$$\text{there is an } h \text{ with } \mathcal{M}, g, h \models \frac{F \cup I \quad I'}{C_{n+1} \dots C_m}$$

A DRT calculus is given in Figure 1 (lists $C_1 \dots C_k$, abbreviated as C). The calculus uses substitution in constraints and DRSs. This notion is defined by

Test axiom	$\frac{\boxed{\begin{array}{ c c } \hline F & \phi \\ \hline C \\ \hline \end{array}}}{\Rightarrow} \boxed{\begin{array}{ c c } \hline F & \phi \\ \hline C \\ \hline \end{array}} \quad M(C) \subseteq F$
Transitivity	$\frac{D \Rightarrow \boxed{\begin{array}{ c c } \hline F & \phi \\ \hline C_1 \\ \hline \end{array}} \quad \boxed{\begin{array}{ c c } \hline F & \phi \\ \hline C_1 \\ \hline \end{array}} \Rightarrow \boxed{\begin{array}{ c c } \hline F & I \\ \hline C_2 \\ \hline \end{array}}}{D \Rightarrow \boxed{\begin{array}{ c c } \hline F & I \\ \hline C_2 \\ \hline \end{array}}}$
Marker intro	$\frac{D \Rightarrow \boxed{\begin{array}{ c c } \hline F & I \\ \hline [t/v] C \\ \hline \end{array}} \quad v \notin F}{D \Rightarrow \boxed{\begin{array}{ c c } \hline F & \{v\} \cup I \\ \hline C \\ \hline \end{array}}}$
Marker shift	$\frac{\boxed{\begin{array}{ c c } \hline F & I \\ \hline C \\ \hline \end{array}} \Rightarrow D}{\boxed{\begin{array}{ c c } \hline F - I_0 & I_0 \cup I \\ \hline C \\ \hline \end{array}} \Rightarrow D}$
Sequencing-l	$\frac{\boxed{\begin{array}{ c c } \hline F & I \\ \hline C \\ \hline \end{array}} \Rightarrow D \quad \boxed{\begin{array}{ c c } \hline F & I \\ \hline C_0 \cup C \\ \hline \end{array}} \Rightarrow D \quad M(C_0) \subseteq F \cup I}{\boxed{\begin{array}{ c c } \hline F & I \\ \hline C_0 \cup C \\ \hline \end{array}} \Rightarrow D}$
Sequencing-r	$\frac{D \Rightarrow \boxed{\begin{array}{ c c } \hline F & I_1 \\ \hline C_1 \\ \hline \end{array}} \quad D \Rightarrow \boxed{\begin{array}{ c c } \hline F & I_2 \\ \hline C_2 \\ \hline \end{array}} \quad I_1 \cap I_2 = \phi}{D \Rightarrow \boxed{\begin{array}{ c c } \hline F & I_1 \cup I_2 \\ \hline C_1 \cup C_2 \\ \hline \end{array}}}$
Neg	$\frac{\boxed{\begin{array}{ c c } \hline F & I \\ \hline C \\ \hline \end{array}} \Rightarrow D \quad \boxed{\begin{array}{ c c } \hline F & I \\ \hline C \cup \{ \neg D \} \\ \hline \end{array}} \Rightarrow \perp \quad \boxed{\begin{array}{ c c } \hline F & I_1 \cup I_2 \\ \hline C_1 \cup C_2 \\ \hline \end{array}} \Rightarrow \perp \quad M(C_1) \cap I_2 = \phi}{\boxed{\begin{array}{ c c } \hline F & I_1 \\ \hline C_1 \\ \hline \end{array}} \Rightarrow \neg \boxed{\begin{array}{ c c } \hline F \cup I_1 & \phi \\ \hline F \cup I_1 & I_2 \\ \hline C_2 \\ \hline \end{array}}}$
Double-neg	$\frac{D \Rightarrow \boxed{\begin{array}{ c c } \hline F & \phi \\ \hline \neg \neg F & I \\ \hline C \\ \hline \end{array}} \quad \boxed{\begin{array}{ c c } \hline F & I_1 \\ \hline C_1 \cup \{ \neg \neg F \cup I_1 & \phi \\ \hline \neg \neg F \cup I_1 & I_2 \\ \hline C_2 \\ \hline \end{array}} \Rightarrow \perp}{D \Rightarrow \boxed{\begin{array}{ c c } \hline F & I \\ \hline C \\ \hline \end{array}} \Rightarrow \perp}$

Figure 1 The calculus for DRT.

$$\begin{aligned}
[t/v]I &:= I \\
[t/v]Pt_1 \cdots t_n &:= P[t/v]t_1 \cdots [t/v]t_n \\
[t/v]\neg D &:= \neg[t/v]D \\
[t/v]\frac{\boxed{\begin{array}{|c|c|} \hline F & I \\ \hline C_1 \cdots C_n \\ \hline \end{array}}}{\vdash} &:= \frac{\boxed{\begin{array}{|c|c|} \hline F & I \\ \hline [t/v]C_1 \cdots [t/v]C_n \\ \hline \end{array}}}{\vdash}
\end{aligned}$$

Of course, when a substitution $[t/v]$ is mentioned in a rule, it is assumed that t is free for v in D . It is also assumed that all DRSs mentioned in the rules satisfy the syntactic well-formedness conditions for DRSs.

Theorem 4

The calculus for DRT is sound.

Proof Induction on the basis that the test axiom is sound and that the rules preserve soundness. Here is one example soundness check, for the rule of marker introduction. Assume $\mathcal{M}, f, g \models D$. Then by the soundness of the premise, there is an h with

$$\mathcal{M}, g, h \models \boxed{\begin{array}{|c|c|} \hline F & I \\ \hline [t/v] C \\ \hline \end{array}}$$

Thus, $\mathcal{M}, h \models [t/v] C$. Since $v \notin F$, h' given by $h'(v) = \llbracket t \rrbracket_b^{\mathcal{M}}$, and $h'(w) = h(w)$ for all $w \neq v$ for which h is defined extends g . By (an appropriate DRT version of) the substitution lemma,

$$\mathcal{M}, g, h' \models \boxed{\begin{array}{|c|c|} \hline F & \{v\} \cup I \\ \hline C \\ \hline \end{array}}$$

This proves

$$D \models \boxed{\begin{array}{|c|c|} \hline F & \{v\} \cup I \\ \hline C \\ \hline \end{array}}$$

Theorem 5

The calculus for DRT is complete.

For the proof of this, we refer to Van Eijck (1999), where the proof system for DRT is related to a proof

system for dynamic predicate logic (Groenendijk and Stokhof, 1991).

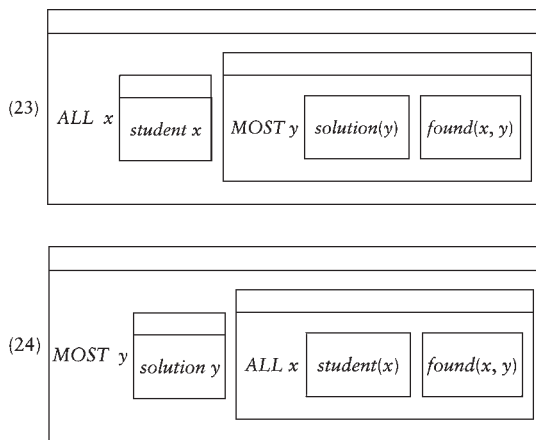
The Treatment of Ambiguities

If an expression of a formal language is viewed as a tree, a partial specification of how the expression is built up from its components can be given by means of a description of constraints on syntax tree construction.

This is the approach to the treatment of ambiguities taken in Underspecified DRT (UDRT) (Reyle, 1993). In UDRT, a DRS is viewed as a tree, and an UDRS is viewed as a set of constraints on tree formation.

(22) *All students found most solutions.*

To represent the scope ambiguity between the two quantifiers in (22), one needs a representation that is ‘in between’ the following two DRSs:



The UDRT solution is to take the DRSs apart, to label the parts, and to define an UDRS as a set of labeled DRS parts plus a list of constraints between labels. An UDRS for the example case has a top node \top labeled l_0 , nodes

$ALL\ x$

$student(x)$

, $MOST\ y$

$solution(y)$

,

$found(x, y)$

labeled l_1, l_2, l_3 , respectively, and constraints $l_0 \leq l_1, l_0 \leq l_2, l_1 \leq l_3, l_2 \leq l_3$. Full disambiguation can be achieved by adding a further constraint. Adding the constraint $l_1 \leq l_2$ disambiguates the UDRS to (23), while adding the constraint $l_2 \leq l_1$ results in disambiguation to (24).

See also: Anaphora: Philosophical Aspects; Context and Common Ground; Coreference: Identity and Similarity; Donkey Sentences; Dynamic Semantics;

Formal Semantics; Montague Semantics; Plurality; Quantifiers: Semantics.

Bibliography

- Asher N (1986). ‘Belief in discourse representation theory.’ *Journal of Philosophical Logic* 15, 127–189.
- Geach P T (1962). *Reference and generality/An examination of some medieval and modern theories* (3rd revised edn., 1980). Ithaca/London: Cornell University Press.
- Groenendijk J & Stokhof M (1991). ‘Dynamic predicate logic.’ *Linguistics and Philosophy* 14, 39–100.
- Heim I R (1982). *The Semantics of Definite and Indefinite Noun Phrases*. Ph. D. diss. University of Massachusetts, Amherst. Published 1987, New York: Garland Press.
- Heim I R (1990). ‘E-Type pronouns and donkey anaphora.’ *Linguistics and Philosophy* 13, 137–177.
- Kamp H (1981). ‘A theory of truth and semantic representation.’ In Groenendijk J, Janssen T & Stokhof M (eds.) *Formal methods in the study of language*. Amsterdam: Mathematisch Centrum. 277–322.
- Kamp H & Reyle U (1990). *From discourse to logic/An introduction to the modeltheoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht: Kluwer.
- Kamp H & Reyle U (1996). ‘A calculus for first order discourse representation structures.’ *Journal of Logic, Language and Information* 5, 297–348.
- Kamp H & Rohrer C (1983). ‘Tense in texts.’ In Bäuerle, Schwarze & Von Stechow (eds.) *Meaning, use and interpretation of language*. Berlin: Walter de Gruyter. 250–269.
- Kripke S (1979). ‘A puzzle about belief.’ In Margalit A (ed.) *Meaning and use*. Dordrecht: Reidel.
- Langholm T (1988). *Partiality, truth and persistence*. CSLI Lecture Notes Number 15, Stanford University (distributed by University of Chicago Press).
- Partee B H (1984). ‘Nominal and temporal anaphora’ *Linguistics and Philosophy* 7, 243–286.
- Reyle U (1993). ‘Dealing with ambiguities by underspecification: Construction, representation and deduction.’ *Journal of Semantics* 10, 123–179.
- Saurer W (1993). ‘A natural deduction system of discourse representation theory.’ *Journal of Philosophical Logic* 22, 249–302.
- Seuren P A M (1985). *Discourse semantics*. Oxford: Basil Blackwell.
- Van Eijck J (1999). ‘Axiomatising dynamic logics for anaphora.’ *Journal of Language and Computation* 1, 103–126.
- Van Eijck J & Kamp H (1997). ‘Representing discourse in context.’ In van Bentham J & ter Meulen A (eds.) *Handbook of Logic and Language*. Amsterdam: Elsevier. 179–237.
- Visser A (1994). ‘Digressions on and desiderata for the design of dynamic discourse denotations.’ Lecture Notes, Department of Philosophy, Utrecht University.

Donkey Sentences

P A M Seuren, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

The problem of ‘donkey sentences’ occupies a prominent place in the logical analysis of natural language sentences. The purpose of logical analysis of sentences is to assign them a structure suitable for logical calculus – that is, the formal derivation of entailments. Some variety of the language of predicate calculus (LPC) is normally used for logical translations.

In LPC, a term in a proposition that has a truth value must either be an expression referring to an individual (or a set of individuals) that actually exists or be a bound variable. Modern predicate calculus is essentially extensional: truth values are computed on the presumption that term referents actually exist, so that it allows in all cases for substitution of coextensional constituents *salva veritate*. Intensional or virtual objects – objects that have merely been thought up but that lack actual existence – have no place in modern logic, just as they have no place in Quine’s ‘desert landscape’ ontology, which has gained currency in large sections of Anglo-Saxon philosophy. That being so, modern logic has no choice but to posit that any argument term of a predicate in a proposition that has a truth value either refers to an actually existing object or is a bound variable ranging over such objects.

Since in natural language one often encounters expressions that have the appearance of being referring argument terms but in actual fact fail to refer – such as the famous sentence in Bertrand Russell’s (1905) article *The present king of France is bald* – Quine (1960) started a “program of elimination of particulars” aimed at reformulating natural language sentences exclusively in terms of the quantificational language of modern predicate calculus, without any referring terms. Thus, for Quine and large sections of the logical community, LPC bans all definite terms and allows only for variables in argument positions.

This, however, will not do for natural language, which has sentences that express purely extensional propositions and yet contain terms that neither refer to an actually existing object nor allow for an analysis as bound variable. These are the so-called donkey sentences. The fact that natural language resists analysis in terms of LPC constitutes the problem posed by the donkey sentences.

The currency of the term ‘donkey sentences’ originates with the British philosopher Peter Geach, whose

discussion of certain sentences, all about donkeys, awakened the interest of modern logicians (Geach, 1962). Geach did not mention – apart from a token reference (1962: 116) to “another sort of medieval example” – that he took his cue from Walter Burleigh (c.1275–after 1344), who introduced donkey sentences in the context of supposition theory, the medieval equivalent of reference theory. In Burleigh (1988: 92), written around 1328, one finds this example:

- (1) *Omnis homo habens asinum videt illum.*
(‘Every man owning a donkey sees it.’)

Burleigh’s problem had nothing to do with LPC, which did not yet exist. His problem was of a different nature. Having noticed that there exist what we now call bound variable pronouns, as in (2), and having stated that these may never take as antecedent a constituent of the same clause (*propositio categorica*), he presented (1) as an apparent counterexample, since the pronoun *illum* takes as antecedent *asinum*, which stands under the same verb (*videt*) and is thus in the same clause.

- (2) All boys expected that the dog would bite *them*.

His answer was that the antecedent of *illum*, i.e., *asinum*, is not a main constituent of the same clause but a constituent of a subordinate predication, i.e., *habens asinum* (‘owning a donkey’).

Geach (1962) discussed the same problem: how to account for the antecedent relation when the antecedent occurs in a relative clause contained in a complex predicate. It stands to reason, he said (1962: 117), to treat *man who owns a donkey* in the sentences (3a) and (3b), which he considered contradictories, as a complex predicate “replaceable by the single word ‘donkey-owner’.” But if we did that, (3a) and (3b) “become unintelligible ... because ‘it’ is deprived of an antecedent”:

- (3a) Any man who owns a donkey beats it.
- (3b) Some man who owns a donkey does not beat it.

A solution could conceivably be found in rewording these sentences as (4a) and (4b) (1962: 117):

- (4a) Any man who owns a donkey, owns a donkey and beats it.
- (4b) Some man who owns a donkey owns a donkey and does not beat it.

Yet, he says, whereas (3a) and (3b) are contradictories, at least according to native speakers’ intuitions, (4a) and (4b) are not (1962: 118):

[F]or both would be true if each donkey-owner had two donkeys and beat only one of them. Medieval logicians would apparently have accepted the alleged equivalences; for they argued that a pair such as [(3a)] and [(3b)] could both be true... and were therefore not contradictories. But plainly [(3a)] and [(3b)], as they would normally be understood, are in fact contradictories; in the case supposed, [(3b)] would be true and [(3a)] false.

The “medieval logicians” Geach argues against are in fact Walter Burleigh, who added the following comment to his discussion of (1), thereby denying that (3a) and (3b) are contradictories (1988: 92–93; translation mine):

It follows that the following are compatible: ‘Every man owning a donkey sees it’ and ‘Some man owning a donkey does not see it’. For assuming that every man owns two donkeys, one of which he sees and one of which he does not see, then it is not only true to say ‘Every man owning a donkey sees it’, but also to say ‘Some man owning a donkey does not see it’. In the same way, suppose that every man who has a son also has two sons, and that he loves the one but hates the other, then both the following are true: ‘Every man who has a son loves him’ and ‘Some man who has a son does not love him’.

Geach’s own solution was to analyze a relative clause within a predicate as an implication under universal, and a conjunction under existential quantification, as in (5).

- (5a) Any man, if he owns a donkey, beats it.
- (5b) Some man owns a donkey and he does not beat it.

This “is quite unforced and does give us a pair of contradictories, as it ought” (Geach, 1988: 92–93). Yet Geach apparently failed to realize that (5a) does not translate into modern predicate logic. Its translation would have to be something like (6), which contains the free variable y in $\text{Beat}(x, y)$

- (6) $\forall x[\text{Man}(x) \rightarrow [\exists y[\text{Donkey}(y) \wedge \text{Own}(x, y)] \rightarrow \text{Beat}(x, y)]]$

Had he realized that, he would have hit on the donkey sentences problem as it lives in modern formal semantics.

Geach strengthened his putative solution by arguing (1972: 115–127) that a sentence like (7) should not be translated as a conjunction of two propositions – as $A \wedge B$ – but rather as a single quantified proposition with *it* translated as a bound variable, as in (8).

- (7) Smith owns a donkey and he beats it.
- (8) $\exists x[\text{Donkey}(x) \wedge \text{Own}(\text{Smith}, x) \wedge \text{Beat}(\text{Smith}, x)]$

His argument amounts to saying that $A \wedge B$ and $A \wedge \neg B$ cannot be true at the same time, whereas (7)

and (9) can. All it takes is for Smith to own two donkeys, only one of which he beats.

- (9) Smith owns a donkey and he does not beat it.

Therefore, Geach argues, the logical translation (8) is correct, since it is compatible with (10), which simply posits a second ass, owned by Smith but not beaten by him

- (10) $\exists x[\text{Donkey}(x) \wedge \text{Own}(\text{Smith}, x) \wedge \neg \text{Beat}(\text{Smith}, x)]$

This analysis, however, cannot be correct, as pointed out in Seuren (2001: 316–318), since it lacks generality in view of cases like (11).

- (11a) Smith must own a donkey, and he may beat it.
- (11b) I believe that Smith owns a donkey, and I fear that he beats it.
- (11c) This made Smith own a donkey and kept him from beating it.

No analysis of the type shown in (8) or (10) is applicable here, since they either require large scope for *a donkey*, which is contrary to what these sentences mean, or have to place the second operator (*may, fear, keep*) in the scope of the first (*must, believe, make*), which again is not what these sentences mean. Geach’s analysis thus comes to nothing.

All this, however, is still beating about the bush. The real problem shows up in (12):

- (12a) Every farmer who owns a donkey feeds it.
- (12b) If Smith owns a donkey, he feeds it.
- (12c) Either Smith does not own a donkey or he feeds it.

In the standard logical analysis of *if* and *or*, they come out as true if Smith owns no donkey. But then *it* cannot be translated as a referring expression (the donkey), as it lacks a referent. It should therefore be translatable as a bound variable. But that, too, turns out to be impossible. Universal quantification, proposed by Quine (1960: 139) and many others as a solution, again falls foul of possible intervening operators, as in (13) (see Seuren, 1998).

- (13a) If Smith wants to own a donkey he must promise to feed it.
- (13b) Either Smith no longer owns a donkey or he still feeds it.

There thus seems to be a hard core of sentences resisting translation into LPC. They contain definite expressions, preferably pronouns, that are neither referring expressions nor bound variables.

Also, these pronouns behave like referring expressions anaphorically linked to an antecedent, and not like bound variable pronouns. The former allow for substitution by a lexical noun phrase (‘epithet anaphora’); the latter do not. Thus, *it* in (14a), (14b),

and (14c) can be replaced by, for example, *the animal*, without much change in meaning, but *it* in (14d), which represents a bound variable, does not allow for such substitution.

- (14a) Smith owns a donkey and he feeds it/the animal.
- (14b) If Smith owns a donkey he feeds it/the animal.
- (14c) Either Smith does not own a donkey or he feeds it/the animal.
- (14d) Every donkey owned by Smith expects that he will feed it/*the animal.

Donkey pronouns, therefore, behave like referring expressions even though they are not allowed to do so under the statutes of current logic. Kamp and Reyle (1993) recognized the fundamental nature of this problem and proposed a radical departure from standard notions and techniques of semantic interpretation. They defended an analysis whereby the donkey pronouns and other definite expressions do not refer directly to entities in the world at hand but instead denote mental representations of possible real-world entities. In this theory, known as *Discourse Representation Theory*, the mechanism of reference is mediated by a cognitive system of mental representations whose relation to any actual world is a matter of independent concern. This halfway station of mental representations creates some extra room for a semantic account of donkey sentences. Even so, however, standard logical analyses are inadequate for natural language. What logic will do better justice to the facts of language is still an open question. Groenendijk and Stokhof (1991) was an attempt at answering that question.

See also: Discourse Representation Theory; Dynamic Semantics.

Bibliography

- Burleigh W (1988). *Von der Reinheit der Kunst der Logik. Erster Traktat. Von den Eigenschaften der Termini.* (De puritate artis logicae. De proprietatibus terminorum). [Translated and edited by Peter Kunze, with introduction and commentary]. Hamburg: Felix Meiner Verlag.
- Geach P T (1962). *Reference and generality: an examination of some medieval and modern theories.* Ithaca, NY: Cornell University Press.
- Geach P T (1972). *Logic Matters.* Oxford: Blackwell.
- Groenendijk J & Stokhof M (1991). 'Dynamic predicate logic.' *Linguistics and Philosophy* 14, 39–100.
- Kamp H & Reyle U (1993). *From discourse to logic: introduction to model-theoretic semantics of natural language, formal logic and discourse representation theory.* Dordrecht: Kluwer.
- Quine W V O (1960). *Word and object.* Cambridge, MA: MIT Press.
- Russell B (1905). 'On denoting.' *Mind* 14, 479–493.
- Seuren P A M (1998). 'Towards a discourse-semantic account of donkey anaphora.' In Botley S & McEnery T (eds.) *New approaches to discourse anaphora: proceedings of the second colloquium on Discourse Anaphora and Anaphor Resolution (DAARC2).* University Centre for Computer Corpus Research on Language, Lancaster University. Technical Papers, Vol. 11, Special Issue. 212–220.
- Seuren P A M (2001). *A view of language.* Oxford: Oxford University Press.

Dthat

D Braun, University of Rochester, Rochester, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

The expression *dthat* (pronounced as a single syllable) was introduced by David Kaplan (1978, 1989b) as a formal surrogate for the English demonstrative *that*. On Kaplan's theory of demonstratives, an utterance of the English demonstrative *that* is typically accompanied by a demonstration that presents an object, namely the demonstratum of the context. The demonstratum is the referent of the demonstrative in the context. The structure of *dthat*-terms reflects this theory of demonstratives. Each *dthat*-term has the form *dthat[α]*, where α is a singular term. Such a term may be used to represent an English demonstrative *cum* demonstration. When so

used, the expression *dthat* serves as a formal surrogate for the demonstrative, while α takes the form of a definite description that serves as a formal surrogate for the demonstration and fixes the reference of *dthat* in every context. In Kaplan's formal possible worlds semantics, the extension of a definite description α , with respect to a context, world, and time, is an individual. The extension of *dthat[α]* with respect to a context, world, and time is the extension of α with respect to the world and time of the context. Thus, *dthat*-terms are rigid designators with respect to a context: Given a context, the extension of a *dthat*-term is the same with respect to all worlds and times. In this respect, *dthat*-terms are like the indexical first-person pronoun *I*, whose extension, given a context, is the same individual (namely the agent of the context) with respect to all worlds and times.

Although Kaplan (1989b) presents his formal semantics in a possible worlds format, he prefers a semantics that uses Russellian structured propositions – propositions whose constituents include individuals, properties, and relations. On such a semantics, the Russellian content of a predicate, with respect to a context, is a property or relation; the content of *I*, with respect to a context, is the agent of the context; and the content of a definite description, with respect to a context, is a complex entity containing the content of the definite description's predicate and the higher-order property or relation expressed by 'the.' Kaplan's (1989b) informal remarks about *dthat* allow for two quite different interpretations of its Russellian semantics (as pointed out in Kaplan, 1989a). On the first interpretation, the Russellian content of a *dthat*-term, in a context, is an individual, namely the referent of α in that context. Kaplan (1989a) calls this the "demonstrative surrogate" interpretation of *dthat*-terms and says that, on this interpretation, the expression *dthat* standing alone is a singular term. But this semantics does not fit well with the syntax that Kaplan (1989b) ascribes to *dthat* in his formal system, where *dthat* appears to be an operator or functional expression. On the second interpretation, the content of a *dthat*-term is a complex entity that contains the content of the

definite description α together with a higher-order content for *dthat*. On this interpretation, *dthat* is a functional expression that rigidifies the attached definite description; it is similar in some respects to the expression *actually* in *the x: actually Ex*. Most theorists who use *dthat*-terms specify that a complete *dthat*-term, *dthat*[α], is a singular term whose content, in a context, is the referent of α in the context.

See also: Deixis and Anaphora: Pragmatic Approaches; Descriptions, Definite and Indefinite: Philosophical Aspects; Indexicality: Philosophical Aspects; Possible Worlds: Philosophical Theories; Propositions; Rigid Designation.

Bibliography

- Almog J, Perry J & Wettstein H (eds.) (1989). *Themes from Kaplan*. Oxford: Oxford University Press.
- Braun D (2001). 'Indexicals.' In Zalta E (ed.) *Stanford encyclopedia of philosophy*. Available at: <http://plato.stanford.edu>.
- Kaplan D (1978). 'Dthat.' In Cole P (ed.) *Syntax and semantics*, vol. 9. New York: Academic Press. 229–253.
- Kaplan D (1989a). 'Afterthoughts.' In Almog *et al.* (eds.) 565–614.
- Kaplan D (1989b). 'Demonstratives.' In Almog *et al.* (eds.) 481–563.

Dynamic Semantics

J Groenendijk and M Stokhof, Universiteit van Amsterdam, Amsterdam, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Information and Information Change

Dynamic semantics is a branch of formal semantics that uses dynamic logic. Dynamic logic was originally developed for the logical analysis of so-called declarative programming languages. The basic idea is that a program can be semantically interpreted as a transition relation between machine states. Two states s and t are related by a program π if, when π is executed in an input state s , t is a possible output state. For example, if the current value of x is 4, the effect of executing the assignment $x := x + 1$ will be that the new value of x is 5. A program is interpreted in terms of the change that it brings about. (Note that for a deterministic program, like the example just given, there is just one possible output state, for a given input state. But if the program threw a die, there would be several possible outputs.)

In standard logical semantics, (indicative) sentences are interpreted in terms of their truth conditions. And substantential expressions are interpreted in terms of their contribution to the truth conditional content of the sentences in which they occur. The truth conditional content of a sentence corresponds to the information it provides about the world. Dynamic semantics takes a different view on meaning. The static notion of information content is replaced by a dynamic notion of information change. The meaning of a sentence is the change in information that it brings about: meaning is 'information change potential.'

Clearly, for a program, a static interpretation in terms of truth conditions makes little sense (as could be argued for imperative or interrogative sentences in natural language). But it is less obvious what the advantages of a dynamic semantics for indicative sentences would be. At first sight, it seems that a static semantics in terms of information content also already makes it possible to give a general characterization of information change potential: 'Add the

information content of the sentence to the information about the world that is already available.'

One general type of argument in favor of a dynamic approach is that, even when we restrict ourselves to purely informative language use, there is more to meaning than can be captured in the notion of meaning as truth conditional content. A standard example (due to Barbara Partee) is the contrast between the following two sequences of sentences:

- (1) I dropped ten marbles and found all of them, except for one. It is probably under the sofa.
- (2) I dropped ten marbles and found only nine of them. ??It is probably under the sofa.

The first sentences in (1) and (2) are truth conditionally equivalent: they provide the same information about the world. Hence, if meaning is identified with truth conditional content, they have the same meaning. However, one may observe that whereas the continuation with the second sentence in (1) is completely unproblematic, the same continuation in (2) is not equally felicitous. The conclusion must be that the two opening sentences differ in meaning; hence, meaning cannot be equated with truth conditional content.

From the viewpoint of dynamic semantics, the two opening sentences in (1) and (2) differ in the way they change information, even though the information they provide about the world is the same. Unlike the opening sentence in (2), the opening sentence in (1) also creates an informational context that licenses the pronoun 'it' in the second sentence.

The notion of context is already important in standard semantics, in the sense that sentences are interpreted relative to a context of utterance, which may include several parameters, such as the time of utterance, etc. So, interpretation depends on context, or on the information available to the speech participants. What dynamic interpretation adds to this is the insight that the process of interpretation also brings about a change of the context, viz., in the information of the speech participants. Both context dependency and the creation of context in speech situations are taken into account.

One of the things this change of perspective brings is that attention shifts from isolated sentences to discourse, or text. It makes it possible to study and describe semantic phenomena that cross the border of a single sentence. As illustrated by the example given above, anaphoric relations across sentences are one such phenomenon.

Discourse Representation Theory and File Change Semantics

Dynamic semantics is not the only logical semantical theory that deals with the incremental interpretation

of discourse, and it was not the first one, either. Whereas the logical roots of dynamic semantics lie in dynamic logic and the semantics of programming languages, its linguistic roots are discourse representation theory and file change semantics.

In discourse representation theory, the interpretation of a discourse – a sequence of sentences – takes the form of an incremental construction of a discourse representation structure. In file change semantics, it is not a single structure that is built, but a system of so-called file cards. The two main elements of discourse representation structures and file card systems are discourse referents and conditions. The discourse referents behave like logical variables, and the conditions are open formulae containing these variables, thereby putting constraints on the possible values of these variables. Discourse referents are introduced by certain noun phrases, in particular indefinites, and correspond to 'individuals talked about' by the discourse. In file change semantics the introduction of a new discourse referent means adding a file card to the system. The conditions are written on the file cards, and can contain information that is spread over several cards, thereby linking them to each other. In discourse representation theory the conditions can be atomic formulae, or more complex structures that include other discourse representation structures.

The discourse referents also function as possible referents for anaphoric expressions. Anaphoric expressions have to be linked to an accessible discourse referent in the discourse representation structure or file card system as it has been constructed at that point. For example, in discourse (1) above, the phrase 'except for one' in the first sentence will have introduced a discourse referent to which the pronoun 'it' in the second sentence can be linked. In discourse (2) the first sentence does not introduce a suitable discourse referent to link 'it' to. So, discourse representation theory and file change semantics can account for the difference in acceptability of the two discourses.

The incremental construction of discourse representation structures is the first step in the interpretation process of a discourse. It clearly has a dynamic nature. Interpreting a sentence may depend on the nature of the structure created by previous sentences, and will result in a change of structure. The second step in the interpretation process of a discourse is more traditional. It consists in a semantic evaluation of the one resulting discourse representation structure relative to a standard logical model. A structure is true in a model if objects from the domain of the model can be assigned to the discourse referents in such a way that all the conditions in the structure are satisfied.

So, although the discourse representation structures are constructed dynamically, their semantic interpretation, and hence, indirectly, the semantic interpretation of the discourses they represent, are interpreted in terms of ‘static’ truth conditional content.

In our discussion of the discourses (1) and (2), we saw that the opening sentences have the same truth conditions. This will also hold for the discourse representation structures they give rise to. We concluded above that since the two discourses as a whole show different semantic behavior, the opening sentences must differ in meaning, and that hence the meaning of sentences cannot be equated with their truth conditions. Discourse representation theory accounts for the difference in meaning, but not in terms of a difference in information *content*, but in terms of a difference in the *representation* of information. The intermediate representation level thus becomes an essential ingredient of the theory of interpretation of natural language.

Dynamic semantics, in particular dynamic predicate logic, was invented to show that the linguistic phenomena that were dealt with in discourse representation theory can also be treated in such a way that the need to postulate an essential intermediate level of representation does not arise. Methodologically and philosophically, this is of importance, since postulating a representational level means to assume a language of thought as an intermediary between language and interpretation. It leads to a mentalistic theory of meaning, inheriting all the philosophical problems that come with such a view. All other things being equal, a theory that allows one to remain neutral on this issue is preferred.

Concerning interpretation, we concentrated on discourse representation theory. The interpretation procedure in file change semantics is different, and it is not so clear that the critique dynamic semantics puts forward against discourse representation theory applies in precisely the same way to file change semantics.

Dynamic Predicate Logic

The typical type of phenomena that discourse representation theory and file change semantics are designed to deal with are exemplified in the following two examples:

(3) A man is walking in the park. He whistles.

(4) Every farmer who owns a donkey beats it.

In standard predicate logic such (sequences of) sentences are a problem. Appropriate translations are as follows:

(5) $\exists x(\text{man}(x) \wedge \text{walk}(x) \wedge \text{whistle}(x))$

(6) $\forall x \forall y ((\text{farmer}(x) \wedge \text{donkey}(y) \wedge \text{own}(x,y)) \rightarrow \text{beat}(x,y))$

The problem with (5) as a translation of (3) is that the translation of the second sentence in (3) has to be brought under the scope of the existential quantifier in the translation of the indefinite in the first sentence. The problem with (6) as a translation of (4) is that the indefinite, which occurs in (4) inside the relative clause in the subject, turns up in (6) as a universal quantifier that scopes over the implication as a whole. Both facts are problematic from a compositional point of view: the logical formulae are composed at odds with the composition of the sentences. From a compositional point of view, we prefer the following formulae as translations:

(7) $\exists x(\text{man}(x) \wedge \text{walk}(x)) \wedge \text{whistle}(x)$

(8) $\forall x((\text{farmer}(x) \wedge \exists y(\text{donkey}(y) \wedge \text{own}(x,y))) \rightarrow \text{beat}(x,y))$

But, under the standard interpretation of predicate logic, these formulae do not express what (3) and (4) express, because the last occurrence of the variable x in (7) and the last occurrence of y in (8) are free occurrences and are outside the scope of the existential quantifiers.

In a nutshell, what dynamic predicate logic achieves is that the translations in (7) and (8) are appropriate translations of (3) and (4), and, in fact, are logically equivalent with (5) and (6), respectively. It succeeds in this by giving a different, dynamic interpretation to the logical constants of the language of predicate logic. The basic feature of the dynamic existential quantifier is that it can bind variables outside its syntactic scope. This is done by interpreting formulae as transition relations between assignments (of values to variables) much in the same way that programs can be viewed as transition relations between machine states (as seen above). That two assignments a and b are related means that when a is an input assignment, b is a possible output assignment. So formulae are interpreted as such input–output relations. Conjunction is interpreted as the composition of the transition relations of the left and the right conjunct. For example, the output of interpreting ‘ $\exists x(\text{man}(x) \wedge \text{walk}(x))$ ’ will consist of assignments that assign an object to x which is both a man and is walking. In interpreting (7) as a whole, these assignments are passed on as input to ‘whistle(x)’ and will remain as output for the whole sequence if the value assigned to x is also an object that whistles.

For (8) the story is slightly more complicated. Possible outputs for ‘ $(\text{farmer}(x) \wedge \exists y(\text{donkey}(y) \wedge \text{own}(x,y)))$ ’ are assignments that assign a particular farmer to x and donkey to y so that that farmer owns y . The effect of processing the implication ‘ $(\text{farmer}(x) \wedge \exists y(\text{donkey}(y) \wedge \text{own}(x,y))) \rightarrow \text{beat}(x,y)$ ’

is that it is checked to see whether every output of the antecedent still has an output after it has served as input for the consequent. In other words, it is checked whether the farmer in question beats all the donkeys that he owns. So, in effect, an existential quantifier inside the antecedent of a conditional amounts to universal quantification over the implication as a whole. Finally, the contribution of the universal quantification over farmers performs the check we just described for every farmer.

The empirical results that dynamic predicate logic obtains for these sentences are precisely the same as those of discourse representation theory and file change semantics, but the tools being used are much more orthodox. Dynamic predicate logic uses the familiar language of predicate logic, and the only difference in the interpretation of the language compared with the standard interpretation is that assignments of values to variables – which are the result of existential quantification – are remembered beyond the syntactic scope of such quantifiers.

Of course, even small changes can have rather big consequences. The logical properties of the system of dynamic predicate logic differ quite radically from those of standard predicate logic.

Update Semantics

Whereas the dynamic semantics of programs and predicate logic were originally formulated in terms of transition relations between states or assignments, update semantics gives a more simple and intuitive way of viewing the dynamics of interpretation. Statements are interpreted in terms of update functions on information states (or contexts, as they are sometimes called). Schematically, the interpretation of a sentence ϕ in an information state s , written as $s[\phi]$, results in a new state t , which now incorporates the information provided by ϕ .

In the simplest case, information states are modeled as sets of possible worlds (as they are used in modal logic). Since our information about the world is usually partial, an information state leaves several possibilities for the way the world could be. If a world w is an element of a state s , this means that w is a way that the actual world could be, according to the information embodied by s . An update $s[\phi]$ will in general lead to a state t such that $t \subseteq s$. Adding information eliminates possibilities. Growth of information makes the information state shrink.

Truth and falsity are not central notions in update semantics. Sentences are not statically evaluated with respect to the world but are interpreted in terms of their effects on information states. Two special cases of such effects are $s[\phi] = \emptyset$ (the empty set), and $s[\phi] = s$.

The former means that updating s with ϕ leads to the *absurd state*, a state in which there are no possibilities left. This typically happens if ϕ is *inconsistent with* the information embodied in s . When $s[\phi] = s$, this means that the update of s with ϕ has no effect. The sentence ϕ provides no new information relative to s . In such cases we say that s *supports* or *accepts* ϕ .

Now, clearly, update semantics gives us an alternative format to spell out the interpretation of a language, as compared to the standard truth conditional approach, but what are the advantages of this alternative view? One way to motivate update semantics is that there are types of sentences that do not obviously have truth conditional content, but that have meanings that can be described in terms of their effects on information states. Perhaps not the best, but certainly the simplest case to illustrate this with, is provided by epistemic modalities such as ‘might.’ Consider the following example:

- (9) It might be raining outside It isn’t raining outside.

The dots indicate that between the utterance of the two sentences some time passes during which, say, the speaker walks to the window to look outside. The point about the example is that this is a consistent, or coherent, piece of discourse. But at the same time, the following is not:

- (10) It isn’t raining outside It might be raining outside.

This would be hard to explain if both sentences were analyzed in terms of truth conditions (for example, using standard modal logic for the analysis of the modality). Since they are composed of the same two sentences, how could the truth conditions of both sequences differ? From an informational dynamic semantic perspective, it is quite clear what makes the difference. Updating the first sentence of (10) leads to an information state in which there are no worlds left where it is raining at this moment. Such an information state would be inconsistent with the second sentence in (10). (The counterfactual ‘It might have been raining’ is all right, of course.) If things are presented in the opposite order, as in (9), we do not meet this problem. As long as our information state leaves open the possibility that it is raining outside, we can happily accept the first sentence in (9). And neither do we need to have a problem in accepting next the information that it isn’t actually raining.

That order matters is a distinctive feature of update semantics, and of dynamic semantics and dynamic logic in general. Also, in the case of anaphoric relations, order is obviously important. The use of an anaphoric expression, such as a pronoun or

anaphoric definite description, is only felicitous in case the preceding discourse has provided a suitable discourse referent to link it to.

Presuppositions

Dynamic semantics is not only suited to give an account of how the information or the context is changed – by eliminating possibilities and adding discourse referents – but also to formulate the conditions that the current information state or context should meet for a sentence to be acceptable at that point in the discourse. Updates in update semantics are typically partial in that they are only defined in states that meet certain requirements. If a sentence has certain presuppositions, it is required that these are already supported or accepted in the current state, and hence do not provide novel information.

Among the different dynamic frameworks, file change semantics most explicitly takes both anaphoric relations and presuppositions into account. Definites, in particular anaphoric definites, are associated with the property of *familiarity*. In terms of the file card metaphor, a definite requires the presence of a card that already carries the information presented in its descriptive content. In contrast, indefinites carry a *novelty* constraint, and should add a new file card to the stock.

The notion of presupposition has a long and complicated history in logical semantics. Without claiming that all puzzles have already been solved, it seems fair to say that the shift from truth conditions to information change and context change potentials has made a substantial contribution to the study of this phenomenon on the border of semantics and pragmatics.

Further Reading

A much more extensive introduction and overview of dynamic semantics and dynamic logic is Muskens, Van Benthem and Visser (1997). Chierchia (1995) provides an introduction in dynamic semantics with emphasis on linguistic motivation. Groenendijk and Stokhof (2000) gives an informal introduction of dynamic semantics, illustrating it with several different kinds of discourse phenomena.

Discourse representation theory was first presented in Kamp (1981). For an extensive overview, see Van Eijck and Kamp (1997); for a full textbook, see Kamp and Reyle (1993). The original sources for file change semantics are Heim (1983, 1989). An important source of inspiration, for dynamic semantics in general, and for Heim in particular, is Stalnaker (1974, 1979).

Dynamic predicate logic was first presented in Groenendijk and Stokhof (1991). A more detailed study of the dynamics of existential quantification can be found in Dekker (1993). A merge of dynamic predicate logic with Montague grammar is given in Groenendijk and Stokhof (1999). A similar merge with discourse representation theory is proposed in Muskens (1995).

The original source for update semantics is Veltman (1996). In Groenendijk, Stokhof and Veltman (1996), the combination of dynamic predicate logic and update semantics is discussed. An extensive overview of the study of presuppositions in dynamic semantics is given in Beaver (1997).

See also: Anaphora: Philosophical Aspects; Definite and Indefinite; Discourse Representation Theory; Donkey Sentences; Montague Semantics.

Bibliography

- Beaver D (1997). 'Presuppositions.' In Van Benthem & Ter Meulen (eds.). 939–1008.
- Chierchia G (1995). *Dynamics of meaning. Anaphora, presupposition, and the theory of grammar*. Chicago: The University of Chicago Press.
- Dekker P (1993). 'Existential disclosure.' *Linguistics and Philosophy* 16, 561–587.
- Groenendijk J A G & Stokhof M J B (1991). 'Dynamic predicate logic.' *Linguistics and Philosophy* 14, 39–100.
- Groenendijk J A G & Stokhof M J B (1999). 'Dynamic Montague grammar.' In Kálmán L & Pólos L (eds.) *Papers from the second symposium on logic and language*. Budapest: Akadémiai Kiadó. 3–48.
- Groenendijk J A G & Stokhof M J B (2000). 'Meaning in motion.' In Von Stechow K & Egli U (eds.) *Reference and anaphoric relations*. Dordrecht: Kluwer. 47–76.
- Groenendijk J A G, Stokhof M J B & Veltman F J M M (1996). 'Coreference and modality' In Lappin S (ed.) *The handbook of contemporary semantic theory*. Oxford: Blackwell. 179–214.
- Heim I (1983). 'File change semantics and the familiarity theory of definiteness.' In Bäuerle R, Schwarze C & Von Stechow A (eds.) *Meaning, use and interpretation of language*. Berlin: De Gruyter. 164–189.
- Heim I (1989). *The semantics of definite and indefinite noun phrases*. New York: Garland.
- Kamp H & Reyle U (1993). *From discourse to logic*. Dordrecht: Reidel.
- Kamp H (1981). 'A theory of truth and semantic representation.' In Groenendijk J A G, Janssen T M V & Stokhof M J B (eds.) *Truth, interpretation and information*. Dordrecht: Foris. 1–41.
- Muskens R (1995). 'Combining Montague semantics and discourse representation.' *Linguistics and Philosophy* 19, 143–186.

-
- Muskens R, VanBenthem J F A K & Visser A (1997). 'Dynamics.' In Van Benthem & Ter Meulen (eds.), 587–648.
- Stalnaker R (1974). 'Pragmatic presuppositions.' In Munitz M & Unger P (eds.) *Semantics and philosophy*. New York: New York University Press. 197–213.
- Stalnaker R (1979). 'Assertion.' In Cole P (ed.) *Syntax and semantics 9: pragmatics*. New York: Academic Press. 315–332.
- VanBenthem J F A K & Ter Meulen A G B (eds.) (1997). *Handbook of logic and language*. Amsterdam: Elsevier.
- VanBenthem J F A K (1991). *Language in action. Categories, lambdas and dynamic logic*. Amsterdam: North Holland.
- VanEijck J & Kamp H (1997). 'Representing discourse in context.' In Van Benthem & Ter Meulen (eds.), 587–648.
- Veltman F J M M (1996). 'Defaults in update semantics.' *Journal of Philosophical Logic* 25, 221–261.
- Zeevat H (1992). 'Presupposition and accommodation in update semantics.' *Journal of Semantics* 9, 379–412.
-

This page intentionally left blank

E

E-Language versus I-Language

G S Rattan, University of Toronto, Toronto, Canada

© 2006 Elsevier Ltd. All rights reserved.

E-Language/I-Language Distinction

The purpose of this article is to clarify the various dimensions of Noam Chomsky's distinction between I- and E-language, with an aim to seeing what is fundamental to that distinction and to discussing some considerations that suggest one or the other conception of language to be correct. The main conclusions are that an E-language conception can and should absorb many of the insights of the I-language conception and that, so enriched, a genuine E-language conception can function foundationally in a theoretical understanding of language.

Some Preliminary Characterizations

Noam Chomsky (1986) introduces the distinction between I-language and E-language with subsequent writings revisiting and clarifying the distinction (e.g., in the papers collected together in Chomsky, 2000). The distinction concerns the metaphysical status of language, of what language is. I-language is internalized, individualistic, idiolectal, and intensional (Chomsky, 2000: 70–73). Each of these features will be discussed in turn, followed by a discussion of the contrasting features of E-language.

Linguistics is to be “conceived of as the study of I-language” and so “becomes part of psychology, ultimately biology” (Chomsky, 1986: 27). To say that a speaker knows a language is to say “the speaker knows an I-language as characterized by the linguist’s grammar” (1986: 27; 40). I-language should be distinguished from Universal Grammar, which we might think of, roughly, as the innately given framework of any human language (*see Innate Knowledge*); I-language is not this initial state of the language faculty but rather “the system of knowledge attained” (1986: 26) at a state of linguistic maturation. I-language is not determined by how speakers use the language, although language use is evidentially relevant for determining the properties of the initial

state and of I-language (cf. George, 1989: note 20). Chomsky also rejects the idea that the language that one speaks is determined by one’s ability with the language (Chomsky, 1980: 51ff.; 1986: 9–21; 2000: 50–52). Rather the language a person speaks is determined by the person’s psychology, in particular by the I-language that is the object of the person’s knowledge. I-language is internalized in being a psychological property of speakers.

I-language is idiolectal in character. The identity conditions for I-languages make no mention of a shared socio-historical object, like English or French, of which speakers have only a partial knowledge. Chomsky (e.g., Chomsky, 2000: 72–73) also rejected Putnam’s (1975) and Burge’s (1979, 1982) antiindividualist ideas that meaning is partially determined by the social and physical environment (*see Externalism about Content*). More will be said about the issue of individualism and antiindividualism below.

Finally, I-language is intensional in the sense that the identity conditions of languages are determined by strong rather than weak equivalence of the grammars that generate them (Chomsky, 2000: 26). Roughly, grammars are weakly equivalent when they generate the same class of expressions as well formed; they are strongly equivalent when they generate the same class in the same way.

I-language is to be contrasted with E-language, which is external, sociolectal, antiindividualist, and extensional. It is external in that the object of E-language study is “a collection of actions, or utterances, or linguistic forms (words, sentences) paired with meanings, or as a system of linguistic forms or events,” perhaps determined by convention (Chomsky, 1986: 19) (*see Conventions in Language*). It is sociolectal in making the primary object of study a social rather than individual-focused object. It is anti-individualist in that it takes the social environment and the physical environment to be relevant to the determination of meaning. And it is extensional in construing languages as identical if the grammar pairs the same sentences with the same meanings, no matter how those sentences or meanings are generated (Quine, 1970).

Getting at What's Fundamental

In order to get at what is fundamental to the distinction between I- and E-language, this article will consider how much of the I-language perspective E-language theorists can accept without compromising their view. As it turns out, the E-language position can accommodate quite a lot. It is noteworthy that the E-language positions described here are either already prominent in the literature or involve natural extensions of the E-language outlook.

Consider first the intensionality of grammar. It is surely part of Chomsky's interest in advocating I-language to rid the study of language of the vestiges of behaviorism; but it is not clear that a quasi-behaviorist conception of language, in which semantic competence consists in dispositions to use sentences under certain conditions, cannot accommodate the intensionality of I-language. Behavioral patterns of semantic acquisition and decay can favor one grammar over some weakly equivalent alternative. If dispositions to use sentences are gained or lost in chunks, that can count in favor of grammars that construe those chunks as wholes grammatically structured in one way rather than another (cf. Evans, 1981 and Davies, 1981b). The same E-language conception shows that the idiolectal aspect of I-language can also be accommodated. The only relevant behavioral dispositions may be that of the individual at a given time. So, a grammar with one structure rather than some other structure weakly equivalent to it can be determined as the grammar of a person at a time. If this is correct, then the intensionality and idiolectal character of language are consistent even with a quasi-behaviorist conception of language and are thus consistent with the overall E-language perspective.

What makes the quasibehaviorist view illegitimate from the I-language perspective? Here other features of I-language are relevant, in particular, the requirement that I-language be internalized. Suppose that the internalization requirement is construed in such a way that in contrast with the quasibehavioral view, that it demands a genuine mentalism; the quasibehavioral conception of I-language is rejected because it fails to be genuinely psychologically realistic. But, again, it seems as though the E-language conception can accommodate I-language requirements. For example, in Lewis's account (1975), language is convention based, and the existence of convention is a psychological matter. Very roughly, in Lewis's account conventions are intentionally propagated regularities among the speech, beliefs, and intentions of speakers and hearers. The point is not the details but rather the fact that conventions are defined partly in terms of the psychological states of individuals. If convention-based conceptions of language are E-language conceptions,

then E-language conceptions of language can accommodate the internalization requirement of I-language when that requirement is interpreted as a requirement to construe language psychologically.

But perhaps there is a stronger way to interpret the internalization requirement. The internalization requirement does not ask merely that language be psychological; it asks that language have a special cognitive home – that language be the object of a special, modularized, faculty of knowledge, the 'language faculty,' that is deployed in speakers' epistemic accomplishments of acquiring the capacity with language and with producing, and perceiving meaning. Language is what is mentally represented in our faculty of linguistic knowledge.

Two aspects of the language faculty should be distinguished. First, the states of the language faculty are tacit and thus differ substantially from ordinary cognitive and conative states such as belief and desire (see **Tacit Knowledge**). Convention-based views of language can accommodate this and perhaps should insist on it. There is at least some *prima facie* case for considering the complicated intentions involved in convention as of a psychologically different kind from the usual, garden-variety intention. Second, and more relevant here, the linguistic faculty is to be a faculty of knowledge of language and thus is a cognitive state directed toward the linguistic facts. Convention-based accounts of language deploy psychological states in the account of convention and may even deploy tacit states, but they do not deploy cognitive states that represent the linguistic facts. That is what the linguistic faculty does. So, the idea is, convention-based accounts of language cannot be I-language accounts because they reject the cognitivism fundamental to the internalization requirement of I-language.

So far we have seen that the intensional and idiolectal, requirements of I-language can be accommodated by E-language conceptions, as can the psychological and tacit aspects of the internalization requirement. But convention-based versions of E-language cannot accommodate the cognitivist aspect of internalization. Can any E-language conception? Again, the answer seems to be affirmative. Indeed, once we move to understanding the internalization requirement as a cognitivist requirement, the E-language conceptions seem to be in its element. States of the language faculty represent the linguistic – the phonological, syntactic, and semantic – facts. Plausibly these facts, like other facts, are facts independent of the thinker. So there seems to be some clear sense in which there is something external about language.

All of this appears to be correct, and it clarifies the issues further. But it doesn't adjudicate the issue between E- and I-language. It fails to do so because

the issue between I- and E-language concerns, not whether the language or the grammatical facts are represented by the language faculty, but rather why some language, rather than some other, is represented by the speaker. The central issue is about the metasemantics of the representations of the language faculty. So even if the facts are external, in the sense that any cognitivism demands, it might well be the case that a speaker stands in a particular representational relation to a language because of facts that are internal to the speaker. The actual language relation may be internally determined.

Individualism and Antiindividualism

This brings us to the final requirement of the I-language conception: that language is not only internalized, but internalist, or individualist, in character. Here we arrive at issues that are fundamental to the debate between I- and E-language. The I-language claim is that language is individualist in the sense that the facts that determine one's language supervene on internal bodily states; the metasemantics of the representation of language is individualist.

According to this kind of individualism, the role of representational content in the states of the language faculty is not questioned; however, it is demanded that the account of why the cognitive states of the language faculty represent what they do appeal only to internal facts. Facts about the language are not identified with facts in the head (for discussion see Rey, 2003a, 2003b) but rather are supervenient upon such facts. However, when this individualist view is applied to semantics, it seems to have the consequence that semantics cannot be referential or truth conditional (*see Truth Conditional Semantics and Meaning*). Representing the referential or truth conditional properties of expressions seems to require standing in some relation to external items of reference, where, in many cases, the external items are external items in the environment. But these relations to external items are not acceptable from an individualist point of view. So it seems that if individualism about semantic cognition is correct, theories of reference and truth can form no part of the content of semantic cognition.

One suggestion for a nontruth-conditional semantics, offered by Ray Jackendoff (1983, 1986), construes the central semantic notion not as truth but as truth-in-a-mental-projection, which is not a replacement for truth but a property that is contrasted with it. In this view, everyday speakers see, speak, and think only of a mentally projected world. This suggestion forces us to deny the instances of the platitudinous T-schema that Tarski (1944) used as a criterion of adequacy for a definition of truth:

(T) *s* is true-in-L iff *p*

where '*s*' is to be replaced by a structural description of a sentence of some language *L*, and '*p*' is to be replaced by the very same sentence (type) in use, or a translation of that sentence. The schema is platitudinous insofar as its use as an adequacy condition is an expression of the platitudinous idea that a sentence is true if and only if things are as use of the sentence says they are. This idea constitutes our fundamental grip on the concept of truth, and it links the concept of truth to the concept of what a sentence says or means (cf. Rattan, 2004) (*see Truth: Theories of in Philosophy*). But in Jackendoff's view, a use of a sentence specifies only a truth-in-a-mental-projection condition, a notion that Jackendoff explicitly contrasts with truth. The view avoids truth conditionalism about meaning, but at too high a cost.

A better suggestion is offered in Pietroski (2003). Pietroski does not deny that language use connects up ultimately with truth (this is its difference from Jackendoff's view). But Pietroski exploits the ambiguity of '*says*' in the platitude about truth to suggest that the platitude is a pragmatic, rather than a semantic, platitude. Indeed, a sentence is true iff things are as a use of the sentence says they are, but in this view, words don't say things, speakers do. Speakers exploit nontruth-conditional properties of expressions to say things, with certain communicative intentions, in certain contexts, that have truth conditions. And, consistently with individualism, the nontruth-conditional properties of expressions are determined by facts inside the head. I-language, including its semantics, is an elaborated, internally determined syntax.

Semantic Intuitions in Linguistics and General Epistemology

This view may ultimately be the most compelling. However, there are grounds for reservation. These grounds center on the role of intuitions in semantics and epistemology (*see Epistemology and Language*).

Intuitions have long been thought to play a fundamental role as data in linguistic theory. These intuitions concern matters of syntax (intuitive judgments of grammaticality or acceptability) and semantics (intuitive judgments of truth conditions). But once the link to truth conditions is broken, the connection between our intuitive judgments about what seems to be a fundamental dimension of what expressions mean – their truth conditions – and semantic theory becomes much more tenuous and theory-involving. Now it may well be that the connection is tenuous and theory-involving, but it would be a mistake to think that this gap simply parallels matters in syntax. In the case of syntax, judgments of grammaticality are overturned based on considerations that have to do with psychological contingencies; for example,

with parsing strategies ('garden-path' sentences) and interactions between memory and syntactic complexity (repeated center-embedding). But if semantics is not truth conditional, then intuitive judgments of truth conditions fail to engage the domain of semantics at all. The failure is not determined by psychological contingencies but rather is preordained by the framework of linguistic theorizing.

Semantic intuitions have another, related role to play in thought. One dimension of Quine's attack (1951) on the analytic/synthetic distinction is on the idea that there is a class of propositions whose truth is grounded in meaning, rather than, as with other true propositions, in worldly fact. Semantic intuitions must then play a dual role in telling us (indirectly) about the language, insofar as the intuitions derive from semantic competence, and also about the world, in being about the world. The second aspect of this dual role can be seen, for example, in the role that semantic intuitions play in correcting lexical explications (Burge, 1986). Considered explications of the meanings of lexical items might determine characteristics that are neither necessary nor sufficient for falling in the extension of, say, *chair*. One traditional, Platonic use of semantic intuitions deploys them to correct mistaken characterizations through the application of lexical items to examples. This deployment is epistemically valuable even when epistemically excessive features often associated with intuitions, like indefeasibility, are stripped away. The general role of semantic intuitions in correcting and being corrected makes sense only against the backdrop of a truth conditional conception of meaning – when there is something that the semantic intuitions intuit or fail to intuit. The epistemic role of semantic intuitions is another casualty of construing semantics in a nontruth-conditional manner.

See also: Conventions in Language; Epistemology and Language; Externalism about Content; Innate Knowledge; Linguistics: Discipline of; Tacit Knowledge; Truth Conditional Semantics and Meaning; Truth: Theories of in Philosophy.

Bibliography

Burge T (1979). 'Individualism and the mental.' *Midwest studies in philosophy* 4, 73–122.

- Burge T (1982). 'Other bodies.' In Woodfield A (ed.) *Thought and object*. Oxford: Oxford University Press. 97–120.
- Burge T (1986). 'Intellectual norms and the foundations of mind.' *Journal of Philosophy* 83, 697–720.
- Chomsky N (1980). *Rules and representations*. New York: Columbia University Press.
- Chomsky N (1986). *Knowledge of language: its nature, origin, and use*. Westport CT: Praeger.
- Chomsky N (2000). *New horizons in the study of language and mind*. Cambridge UK: Cambridge University Press.
- Davies M (1981). *Meaning, quantification and necessity: themes in philosophical logic*. London: Routledge and Kegan Paul.
- Evans G (1985 [1981]). 'Semantic theory and tacit knowledge.' In *Collected papers*. Oxford: Oxford University Press. 322–342.
- George A (1989). 'How not to become confused about linguistics.' In George A (ed.) *Reflections on Chomsky*. Oxford: Blackwell. 90–110.
- George A (1990). 'Whose language is it anyway: some notes on idiolects.' *Philosophical Quarterly* 40, 275–298.
- Jackendoff R (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff R (1986). *Semantic structures*. Cambridge, MA: MIT Press.
- Lewis D (1983 [1975]). 'Languages and language.' Reprinted in his *Philosophical papers: Volume 1*. Oxford: Oxford University Press. 163–188.
- Pietroski P (2003). 'The character of natural language semantics.' In Barber A (ed.) *Epistemology of language*. Oxford: Oxford University Press. 217–256.
- Putnam H (1975 [1975]). 'The meaning of "meaning."' Reprinted in his *Mind, language, and reality: philosophical papers volume 2*. Cambridge: Cambridge University Press. 215–271.
- Quine W V O (1963 [1951]). 'Two dogmas of empiricism.' Reprinted in his *From a logical point of view*. Cambridge: Harvard University Press. 20–46.
- Quine W V O (1970). 'Methodological reflections on current linguistic theory.' *Synthese* 21, 386–398.
- Rattan G S (2004). 'The theory of truth in the theory of meaning.' *European Journal of Philosophy* 12, 214–243.
- Rey G (2003a). 'Chomsky, intentionality and a CRTT.' In Antony L & Hornstein N (eds.) *Chomsky and his critics*. Oxford: Blackwell. 105–139.
- Rey G (2003b). 'Intentional content and a Chomskyan linguistics.' In Barber A (ed.) *Epistemology of language*. Oxford: Oxford University Press. 140–186.
- Tarski A (1944). 'The semantic conception of truth and the foundations of semantics.' *Philosophy and Phenomenological Research* 4, 341–376.

Empiricism

J Prinz, University of North Carolina, Chapel Hill, NC, USA

© 2006 Elsevier Ltd. All rights reserved.

The term ‘empiricism’ refers to a group of related philosophical doctrines, which place special emphasis on perception or experience. On one traditional formulation, empiricists hold that there is nothing in the intellect that is not first in the senses. This idea has roots in the texts of Aristotle and other ancient philosophers, but it was most influentially developed and defended by the British empiricists of the 17th and 18th centuries (see Locke, 1690; Hume, 1739). Empiricism took on new forms in the twentieth century. The history of Western philosophy can be usefully viewed as an extended debate between empiricists and rationalists.

Species of Empiricism

Empiricists say that experience is primary. But primary in what sense? There are several possible answers: primary as a source of knowledge; primary in the makeup of mental concepts; and primary in linguistic meaning. Each of these answers corresponds to a distinct species of empiricism. This article explains all three in turn.

In epistemology, empiricism is the doctrine that knowledge is based on experience (or ‘*a posteriori*’). Some epistemological empiricists claim that all knowledge is based on experience, some claim that most knowledge is, and some merely insist that the paradigm cases of knowledge are based on experience.

Historically, epistemological empiricism was contrasted with epistemological rationalism: the view that knowledge derives from intuition or pure reason. Rationalists claim that knowledge can be derived prior to experience (or ‘*a priori*’), and, to make good on this claim, they typically assume that we have an extensive stock of innate ideas. In ancient Greek philosophy, Aristotle and Plato inaugurated the debate between epistemological empiricists and rationalisms. Aristotle emphasized the role of observation in learning, whereas Plato emphasized innate knowledge that is acquired prior to birth and either triggered by experience (in the way that a memory might be triggered by seeing a related cue) or else revealed through philosophical reflection. In modern British philosophy, epistemological empiricism was resuscitated by Francis Bacon. Scientific method, he argued, depends on testing hypotheses against experience. Some empiricists have been willing to grant that certain domains are known *a priori*. John Locke

holds this view of ethics and mathematics. Other empiricists have been more thoroughgoing. Moral sense theorists, such as Francis Hutcheson, postulate a special faculty that uses the passions – a kind of sense – to detect ethical truths. Empiricists about mathematics claim that we obtain mathematic knowledge by observation. It has been claimed that we can observe basic truths of geometry, arithmetic, and even set theory.

Epistemological empiricists typically endorse both a causal thesis and a normative thesis. According to the causal thesis, knowledge is causally obtained through experience. For example, Hume says that we obtain knowledge by association: We associate ideas of things that are experienced in spatial or temporal proximity. Associationism exerted an influence on behaviorism and connectionism. Behaviorists say we obtain knowledge by conditioning, which occurs when a stimulus or emitted behavior is accompanied by a behavioral reinforcer. Connectionists believe that knowledge acquisition involves the modulation of weighted connections between layered populations of neurons.

The normative thesis endorsed by epistemological empiricists says that experience is the primary source of epistemic justification. This thesis has often taken the form of epistemic foundationalism: the idea that knowledge must always be justified with reference to things that can be known directly. According to some foundationalists, the only things that we know directly are sense data: perceptual experiences given to us when our sense organs are stimulated (Ayer, 1936). Knowledge that goes beyond direct experience must be based on sense data. For example, Bertrand Russell (1912) argued that we obtain knowledge of the external world by inductive inference: We postulate a mind independent reality because it offers the best explanation of the fact that sense data seem to be outside our control. We might experience the shape of a cat, and then turn away, only to find that the shape remains when we return our gaze to the original position. The best explanation of this is that there is a cat in that location and it exists even when we are not viewing it. Some empiricists, who grant that all knowledge must be sense-based, find this argument unpersuasive; they are driven to skepticism about the external world or to the view that the external world is nothing but a collection of ideas (a view known as idealism or phenomenalism).

Epistemological empiricists need not be foundationalists. W. V. O. Quine (1951) claimed that theories must be confirmed against experience in entirety, rather than confirming a privileged class of

observational judgments and deriving the rest from those. Quine's alternative to foundationalism is called confirmation holism. On this approach, a judgment gains its epistemic warrant from experience along with its coherence with all the other judgments that have been supported by experience and pragmatic rules guiding theory revision. William James and John Dewey also defended versions of epistemological empiricism without foundationalism.

Epistemological empiricism can be distinguished from concept empiricism, the view that all of our concepts (or 'ideas') are derived from percepts (or 'impressions'). Ideas are encoded in the representational formats that are used by the senses. Aristotle said there is no idea without an image. John Locke and David Hume argued that ideas are abstracted or copied from impressions. This doctrine has had a recent revival in cognitive science (see Barsalou, 1999; Prinz, 2002). In addition to our external senses, concept empiricists admit that we can form ideas on the basis of 'reflection,' or inner experiences, such as the passions and sentiments. For example, Hume tried to explicate the concept of moral goodness by appeal to the sentiment of approbation.

Concept empiricists typically reject the rationalists' contention that many ideas are innate. Instead, ideas derive from 'impressions' (the mental states caused in our senses when we perceive the world). According to Locke, some ideas are simply stored copies of impressions, some are combinations of impressions, and some are derived from impressions through abstraction: a process (or set of processes) by which we extract general ideas from experiences of particulars.

In addition to epistemological and concept empiricism, there is an empiricist doctrine relating to meaning. Semantic empiricists contend that the meaning of every linguistic expression must be specifiable using expressions that ultimately get their meaning from experience. One version of this view is the principle of verifiability, which was defended by the logical positivists in the beginning of the twentieth century. According to this principle, a sentence is meaningful only if it can be confirmed or falsified by experience. All meaningful sentences use expressions that refer to things that can be directly observed (observation terms) or things that can be inferred from observable things (theoretical terms).

They allowed one kind of exception, however. The positivists drew a distinction between sentences that can be confirmed by experience (synthetic truths) and sentences that are true in virtue of the meanings of their terms (analytic truths). The sentence 'all mothers are female' is analytic, because 'mother' means female parent. The sentence 'Madonna is a mother' is synthetic. Analytic truths are, in a sense,

knowable *a priori*, but they need not violate the spirit of epistemological empiricism. Analytic truths are based on relations between words that are stipulated by the authors of a language. To accept *a priori* knowledge of such truths is not a concession to the rationalists' claim that we have *a priori* access to substantive truths about the world. Nevertheless some epistemological empiricists reject analytic truths. Quine famously argued that no truth is immune from empirical refutation. New discoveries or events can lead us to revise apparent tautologies. If cross-dressing men began to adopt children and conform to prototypical maternal roles, then we might begin to refer to them as 'mothers.' Thus, 'all mothers are female' would be refuted.

These forms of empiricism often have been defended as a package, but they are independent. For example, an epistemological empiricist might reject concept empiricism and instead adopt the view that concepts are couched in a nonsensory language of thought. A concept empiricist might reject epistemological empiricism and argue for a coherentist theory of justification instead of foundationalism. Concept empiricists and epistemological empiricists might replace semantic empiricism with a causal theory of meaning, which would allow reference to unobservable entities.

Empiricism and Language

Each empiricist doctrine has implications for the study of language. This is most obvious in the case of semantic empiricism, which is a thesis about linguistic meaning. The link between epistemological empiricism and language is most remote, because theories of justification are usually held to be independent of theories of meaning. For verificationists, however, these two dimensions collapse, because meanings are equated with conditions for obtaining knowledge. Some verificationists were led by epistemological considerations (the impossibility of confirming entities hidden from view) to their thesis about meaning.

The connection between concept empiricism and language depends on the relationship between concepts and meanings. Some authors believe that concepts are private mental representations and that meanings are either mind-independent properties and propositions or public rules of linguistic behavior. Other authors believe that meanings are at least partially constituted by the concepts in our heads. Those in this latter camp would claim that an empiricist theory of concepts entails a form of empiricism about semantics. Concept empiricism would entail that meanings are at least partially constituted by perceptually derived mental representations.

Some cognitive scientists try to bridge concept empiricism and semantic empiricism by collecting evidence that people spontaneously use perceptual representations when understanding words and phrases (see, e.g., Barsalou, 1999). A related synthesis can be found in the work of some cognitive grammarians. They argue that some meanings are based on conceptual knowledge and that some concepts are partially constituted by image schemas. Image schemas are schematic (i.e., nonmetric) representations of spatial relations that are used in perceiving the world. George Lakoff (1987) has argued that highly abstract concepts, which are believed to pose a challenge for empiricists, are understood by means of metaphorical extension from perceivable spatial relations. The concept of ownership, for example, is understood with reference to the relation of spatial containment.

Some concept empiricists explain our mastery of abstract concepts in a different way. They argue that public language can be used as vehicles of thought. Sentences are observable. If concepts are stored records of experiences, they can be stored records of experiences with public linguistic items. To understand an abstract concept, on this approach, is to master a set of verbal entailments. Benjamin Whorf and Edward Sapir may have been led to their hypothesis of linguistic relativity in virtue of accepting a picture like this.

Another topic of concern to empiricists is language acquisition. Because empiricists traditionally reject innate ideas, some have argued that language can be acquired using general-purpose perceptual learning rules (such as pattern recognition, association, and conditioning). This view is highly controversial.

Noam Chomsky and his followers have developed powerful arguments for the conclusion that language acquisition requires innate learning mechanisms that are specifically designed for language. Chomsky is a self-proclaimed rationalist, and some empiricists hope to prove that his arguments are mistaken.

In sum, empiricism is a family of doctrines united by the central role they afford to experience. These doctrines often have been defended by the same authors, but they are actually independent. Each has several forms, each faces different challenges, and each has implications for the nature of language.

See also: Behaviorism: Varieties.

Bibliography

- Ayer A J (1936). *Language, truth, and logic*. New York: Dover.
- Barsalou L W (1999). 'Perceptual symbol systems.' *Behavioral & Brain Sciences* 22, 577–660.
- Hume D (1739/1978). Nidditch P H (ed.) *A treatise of human nature*. Oxford: Oxford University Press.
- Lakoff G (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
- Locke J (1690/1979). *An essay concerning human understanding*. Nidditch P H (ed.) Oxford: Oxford University Press.
- Prinz J J (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: MIT Press.
- Quine W V O (1951/1953). 'Two dogmas of empiricism.' Reprinted in *From a logical point of view*. Cambridge, MA: Harvard University Press.
- Russell B (1912). *The problems of philosophy*. New York: Henry Holt and Company.

Empty Names

B Caplan, University of Manitoba, Winnipeg, Canada

© 2006 Elsevier Ltd. All rights reserved.

Most names refer. For example, the name 'Uma Thurman' refers to Uma Thurman. But some names don't refer. For example, suppose that Keanu Reeves introduces the name 'Sparkie' to refer to the lighter in his pocket, if there is one, and to nothing otherwise. If it turns out that Keanu's pockets are empty, then 'Sparkie' doesn't refer. In that case, 'Sparkie' is an *empty name*. Usually, an empty name is empty because there is no object for it to refer to. This is

the case with 'Sparkie': 'Sparkie' is empty because Keanu's pockets are empty. But sometimes an empty name is empty not because there is no object for it to refer to but rather because of something else: the speaker's intentions, say. For example, some philosophers think that although fictional characters exist, names from fiction are empty when they are used with the intention of telling a story. On this view, there is a fictional character that *we* can use the name 'Sherlock Holmes' to refer to when we intend to talk about Sir Arthur Conan Doyle's work, but when *Conan Doyle* uses that name to tell a story, it's empty.

The term 'empty name' goes back at least to the German mathematician and philosopher Gottlob Frege (1848–1925), who called expressions that don't refer 'empty' (or, in German, *leer*). The problems that empty names pose go back much further, at least to the Greek philosopher Parmenides (5th century B.C.E.), who argued that you cannot say of what does not exist that it does not exist. Empty names have been giving philosophers headaches ever since.

The Problem of Negative Existentials

Consider a subject-predicate sentence of the form $[\alpha \varphi]$, where α is the subject and φ is the predicate. $[\alpha \varphi]$ says, of the object that α refers to, that it has the property that φ refers to. And $[\alpha \varphi]$ is true if and only if what it says is the case: that is, if and only if that object has that property. For example, (1) says, of the object (or, in this case, person) that 'Uma' refers to, that it has the property that 'is a movie star' refers to.

(1) Uma is a movie star.

'Uma' refers to Uma, and 'is a movie star' refers to the property *being a movie star*. So (1) says, of Uma, that she has the property *being a movie star*. Since Uma has that property, (1) is true. Now consider (2), which says, of the object that 'Sparkie' refers to, that it has the property that 'doesn't exist' refers to.

(2) Sparkie doesn't exist.

The predicate 'doesn't exist' refers to the property *not existing* (or *being nonexistent*). (2) is a *negative existential*, since it appears to say of some object that it doesn't exist. The trouble with negative existentials is that some of them seem true; but it's hard to see how any of them could be. For example, (2) seems true. But since 'Sparkie' is empty, there is no object that it refers to; so there is no object for (2) to say, of it, that it has the property *not existing*. As a result, it is hard to see what, if anything, (2) says. And if (2) doesn't say anything, then it is hard to see how it could be true.

More generally, negative existentials are a problem for anyone who thinks that (a) the truth value of a sentence is determined, compositionally, on the basis of some semantic feature or features of its parts, and (b) the relevant semantic feature of a name is its referent. (And many philosophers share these assumptions: for example, Fregeans, Millians, and Davidsonians all can accept [a] and [b].) Since an empty name has no referent, it seems to follow that negative existentials that contain empty names have no truth value. This is a problem, because such negative existentials seem true.

Millianism

Empty names pose a number of problems in particular for Millianism. According to Millianism, sentences express *propositions*. These are abstract objects that are the primary bearers of truth values. They are also the objects of attitudes, such as believing and asserting. According to Millianism, propositions are *structured*. This means that the proposition expressed by (1), for example, can be represented as the ordered pair $\langle U, \text{being a movie star} \rangle$, where U is something that corresponds to 'Uma.' (In what follows, complications about what goes in the non-U slot are glossed over; it is assumed that what goes in that slot is a property.) And, according to Millianism, sentences that contain names express *singular* propositions if they express any propositions at all. This means that, in the proposition expressed by (1), U is Uma herself. The proposition expressed by (1) can then be represented as $\langle \text{Uma}, \text{being a movie star} \rangle$.

The source of the problems that empty names pose for Millianism is that it seems that Millianism entails that a sentence that contains an empty name doesn't express any proposition at all. According to Millianism, propositions are structured; so, if (2) expresses a proposition, then that proposition can be represented as $\langle S, \text{not existing} \rangle$, where S is something that corresponds to 'Sparkie.' And, according to Millianism, sentences that contain names express singular propositions, if they express any propositions at all; so, if (2) expresses a proposition that can be represented as $\langle S, \text{not existing} \rangle$, then S is the object that 'Sparkie' refers to. But, since 'Sparkie' is empty, there is no object that it refers to. So there is no object in the S slot in $\langle S, \text{not existing} \rangle$. As a result, it seems that there is no singular proposition for (2) to express and hence that, according to Millianism, (2) doesn't express any proposition at all. The view that sentences that contain empty names don't express any proposition at all is called the *No Proposition View*.

The No Proposition View apparently has a number of consequences that are apparently counterintuitive. For example, you might think that a sentence is meaningful only if it expresses a proposition. If that's right, then the No Proposition View entails that (2), for example, is meaningless. But (2) doesn't seem meaningless. This problem is called the *Problem of Meaningfulness*. Or you might think that a sentence inherits its truth value from the proposition it expresses. If that's right, then the No Proposition View entails that, for example, (2) and (3) have no truth value.

(3) Keanu believes that Sparkie doesn't exist.

But (2) and (3) seem true. This problem is called the *Problem of Truth Value*. Or you might think that a

person can sincerely and assertively utter a sentence only if she believes the proposition that it expresses. If that's right, then the No Proposition View entails that no one can sincerely and assertively utter (2), for example. But it seems that someone could sincerely and assertively utter (2). This problem is called the *Problem of Belief and Sincere Assertive Utterance*. Because of these problems (and others that have to do with the substitution of coreferential names in various linguistic contexts), many philosophers have concluded that Millianism should be rejected in favor of its rival, Fregeanism.

Fregeanism

Fregeanism can solve many of the problems that empty names pose for Millianism. Fregeanism agrees with Millianism that sentences express structured propositions. According to Fregeanism and Millianism alike, the proposition expressed by (1) can be represented as $\langle U, \textit{being a movie star} \rangle$ (again, glossing over what goes in the non-U slot). But Fregeanism denies that sentences that contain names express singular propositions. According to Fregeanism, in the proposition expressed by (1), U is *not* Uma herself. Rather, U is a *mode of presentation* of Uma, MP_{Uma} , something that is a way of thinking about Uma or that captures an agent's perspective on Uma. Perhaps MP_{Uma} is something that picks Uma out by describing her as having certain properties: *being the lead in Kill Bill Vol. 1*, say. Or perhaps MP_{Uma} is some other kind of entity. Where Millians say that the proposition represented as $\langle \textit{Uma}, \textit{being a movie star} \rangle$ is true if and only if Uma has the property *being a movie star*, Fregeans say that the proposition represented as $\langle MP_{\text{Uma}}, \textit{being a movie star} \rangle$ is true if and only if the object that MP_{Uma} presents has the property *being a movie star*. Since MP_{Uma} presents Uma, Fregeans agree with Millians that the proposition that (1) expresses is true if and only if Uma has the property *being a movie star*.

Fregeans can reject the No Proposition View. According to Fregeanism, the proposition expressed by (2) can be represented as $\langle S, \textit{not existing} \rangle$. But, according to Fregeanism, S isn't the object that 'Sparkie' refers to; rather, S is a mode of presentation, MP_{Sparkie} , that corresponds to 'Sparkie.' (If there is such a mode of presentation, then there can be empty modes of presentation: that is, modes of presentation that don't actually present anything. But see **Object-Dependent Thoughts**.) Perhaps MP_{Sparkie} has something to do with the property *being a lighter in Keanu's pocket* and would pick out the unique object that has that property, if there were such an object. Or perhaps MP_{Sparkie} is some other kind of entity.

Fregeans can solve the Problem of Meaningfulness: (2) is meaningful because it expresses the proposition represented as $\langle MP_{\text{Sparkie}}, \textit{not existing} \rangle$. Fregeans can also solve the Problem of Truth Value, at least for belief ascriptions like (3). (3) can be true because Keanu can believe the proposition represented as $\langle MP_{\text{Sparkie}}, \textit{not existing} \rangle$. (How the truth of [3] can be compatible with the assumptions [a] and [b] mentioned earlier is complicated. Perhaps in [3] 'Sparkie' isn't really empty: perhaps in [3] Sparkie refers to MP_{Sparkie} . Or perhaps in [3] 'that' refers to the proposition that [2] expresses.) And Fregeans can solve the Problem of Belief and Sincere Assertive Utterance: speakers can sincerely and assertively utter (2) because they can believe the proposition represented as $\langle MP_{\text{Sparkie}}, \textit{not existing} \rangle$.

But Fregeanism doesn't straightforwardly solve the Problem of Truth Value for sentences like (2). For (2) is true if and only if the proposition that it expresses, the proposition represented as $\langle MP_{\text{Sparkie}}, \textit{not existing} \rangle$, is true; and that proposition is true if and only if the object that MP_{Sparkie} presents has the property *not existing*. But there is no object that MP_{Sparkie} presents, and hence it is not the case that the object that MP_{Sparkie} presents has the property *not existing*. So (2) isn't true. Still, many philosophers think that overall Fregeanism fares better than Millianism in handling the problems that empty names pose.

More Millianism: The Gappy Proposition View

Some Millians reject the No Proposition View in favor of the *Gappy Proposition View*, according to which (2) expresses a *gappy proposition* that can be represented as $\langle _, \textit{not existing} \rangle$. (That proposition is gappy because it is just like a singular proposition except that it contains no object where a singular proposition would.) The Gappy Proposition View can solve the Problem of Meaningfulness: (2) is meaningful because it expresses the gappy proposition represented as $\langle _, \textit{not existing} \rangle$. The Gappy Proposition View can also solve the Problem of Truth Value, at least for belief ascriptions like (3). (3) can be true because Keanu can believe the proposition represented as $\langle _, \textit{not existing} \rangle$. And the Gappy Proposition View can solve the Problem of Belief and Sincere Assertive Utterance: speakers can sincerely and assertively utter (2) because they can believe the proposition represented as $\langle _, \textit{not existing} \rangle$.

But the Gappy Proposition View can't solve the Problem of Truth Value in general. Suppose that Uma introduces 'Markie' to refer to the pen in her pocket, if there is one, and to nothing otherwise. If it

turns out that Uma's pockets are empty, too, then 'Markie' doesn't refer either. On the Gappy Proposition View, (4) and (5) express the same gappy proposition, which is a conditional whose antecedent can be represented as $\langle ___, \text{existing} \rangle$.

- (4) If Sparkie exists, then there is a lighter in Keanu's pocket.
- (5) If Markie exists, then there is a lighter in Keanu's pocket.

But (4) and (5) seems to differ in truth value: (4) seems true, whereas (5) doesn't.

Millians could co-opt some of the resources of Fregeanism and say that agents believe propositions via modes of presentation. (4) and (5) express the same gappy proposition; but there are different modes of presentation associated with (4) and (5). The mode of presentation associated with (4) has something to do with $MP_{\langle \text{Sparkie} \rangle}$, whereas the mode of presentation associated with (5) has something to do with $MP_{\langle \text{Markie} \rangle}$. On this view, (4) seems true, because agents believe, via the mode of presentation associated with (4), the gappy proposition expressed by (4) and (5); but (5) doesn't seem true, because agents don't believe that proposition via the mode of presentation associated with (5). Let's call a Millian view that co-opts Fregean resources in some way *Neo-Millian*.

Still More Millianism: The Communicated Proposition View

Once Millians co-opt Fregean resources and become Neo-Millians, they no longer have to appeal to gappy propositions. Neo-Millians can say that *sentences* that contain empty names don't *express* propositions; but when they use those sentences, *speakers* can *communicate* propositions and in fact speakers can communicate the very propositions that Fregeans say are expressed by the sentences that speakers use. Let's call this view the *Communicated Proposition View*. The Communicated Proposition View can solve the Problem of Meaningfulness: (2) is meaningful because speakers can use it to communicate the proposition represented as $\langle MP_{\langle \text{Sparkie} \rangle}, \text{not existing} \rangle$. The Communicated Proposition View can also solve the Problem of Truth Value: (2) seems true because speakers use it to communicate the proposition represented as $\langle MP_{\langle \text{Sparkie} \rangle}, \text{not existing} \rangle$; and (3) seems true because speakers use it to communicate that Keanu believes that proposition. And the Communicated Proposition View can solve the Problem of Belief and Sincere Assertive Utterance: speakers can sincerely and assertively utter (2)

because they can believe the proposition represented as $\langle MP_{\langle \text{Sparkie} \rangle}, \text{not existing} \rangle$.

The possibility of Neo-Millianism suggests that the debate between Millians (including Neo-Millians) and Fregeans should be understood not as a debate about *whether* there are modes of presentations like $MP_{\langle \text{Sparkie} \rangle}$ but rather as a debate about *where* there are such modes of presentation. Fregeans say that such modes of presentation are constituents of the propositions expressed by sentences that contain names. Millians, by contrast, say that they're not, although they might mediate agents' cognitive relations to propositions (as on the Neo-Millian version of the Gappy Proposition View) or they might be constituents of propositions that speakers communicate (as on the Communicated Proposition View).

See also: Assertion; Causal Theories of Reference and Meaning; Communication, Understanding, and Interpretation: Philosophical Aspects; Compositionality: Philosophical Aspects; Direct Reference; Existence; Fictional Discourse: Philosophical Aspects; Object-Dependent Thoughts; Proper Names: Philosophical Aspects; Propositional Attitude Ascription: Philosophical Aspects; Propositions; Reference: Philosophical Theories; Rigid Designation; Semantic Value; Semantics–Pragmatics Boundary; Sense and Reference: Philosophical Aspects; Truth: Primary Bearers.

Bibliography

- Braun D (1993). 'Empty names.' *Noûs* 27(4), 449–469.
- Braun D (in press). 'Empty names, mythical names, fictional names.' *Noûs*.
- Cartwright R (1960). 'Negative existentials.' *Journal of Philosophy* 57(20–21), 629–639. [Reprinted in Cartwright R (1987). *Philosophical essays*. Cambridge, MA: MIT. 21–31.]
- Donnellan K S (1974). 'Speaking of nothing.' *Philosophical Review* 83(1), 3–31.
- Evans G (1982). McDowell J (ed.). *The varieties of reference*. Oxford: Clarendon.
- Frege G (1892). 'Über Sinn und Bedeutung' ["On sense and reference"]. *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50. [Also, 'Frege G (1952 [1980]) "On sense and meaning".' In Black M (trans.), Geach P T & Black M (eds.) *Translations from the philosophical writings of Gottlob Frege*, 3rd edn. Oxford: Blackwell. 56–78. Also, Frege G (1997). 'On *Sinn* and *Bedeutung*.' In Beaney M (ed.) *The Frege reader*. Oxford: Blackwell 151–171.]
- Kripke S (1973). Reference and existence: the John Locke lectures for 1973. Manuscript. Cited in Evans (1982) and Salmon (1998).
- Salmon N (1998). 'Nonexistence.' *Noûs* 32(3), 277–319.

Epistemology and Language

H Jackman, York University, Toronto, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

There has always been a close relation between theories of language, mind, and knowledge, and over the past 80 years, the connection between epistemology and language has typically manifested itself in one of two ways.

First of all, reflection on language has been used to give an account of the meaning of epistemic terms like 'know,' 'rational,' and 'justified,' which have nontrivial, often antiskeptical, consequences. For instance, Strawson (1952) argued that using induction is an essential part of what we *mean* by having 'reasonable' or 'justified' beliefs about the world around us, and so it made no sense to worry about the 'rationality' or 'justification' of induction. More counterintuitively, Malcolm (1952) argued there was a "strong" sense of "know" in which it was impossible for him to doubt that he knew that there was, say, a pen in his hand, since he would have counted *nothing* as evidence against this claim.

This sort of 'ordinary language' approach has generally (and justly) fallen out of favor, as its presuppositions about the meaning of epistemic terms are now viewed as implausible. However, a related strategy that drew from the same Wittgensteinian roots (Wittgenstein, 1969) has emerged in the form of contemporary 'contextualist' epistemologies. These argue that our use of 'know' and related terms is context sensitive in ways that allow the skeptic's claim that we know nothing about, say, the external world *and* our everyday claims to know commonplace facts about the world to *both* be true. It's just that the standards for applying the term 'knowledge' (like those for the application of terms like 'tall') shift from context to context (DeRose, 1995; Williams, 1991).

Such 'therapeutic' approaches do not so much solve traditional skeptical problems as show them to rely upon misunderstandings about how our epistemic terms operate. There is, however, a second way in which philosophical theories of language can be used to respond more 'directly' to traditional epistemological problems.

The second approach focuses not upon the meaning of specific terms, such as 'know,' but rather upon our ability to derive epistemological conclusions from any *general* account of meaning that outlines the limits of our representational capacities. The limits of our representational capacities determine the extent to which certain sorts of skeptical doubts are

legitimate and a general theory of language could thus be used as a weapon against skepticism.

To take an illustrative example, classical verificationism, which understood the meanings of our utterances in terms of the experiences taken to verify them, had clear antiskeptical consequences. If the meaning of a statement like 'there is a pen in front of me' were determined by the experiences standardly used to confirm it, and I could *know* that I'm having such experiences, then I could know that there is a pen in front of me. The verificationist theory of meaning thus blocked skepticism regarding the transition from knowledge of our experience to knowledge of the 'external' world (Ayer, 1936). Of course, it did this at the expense of making problematic our being able to talk or think about anything other than our own experiences.

This general pattern carries over to most other attempts to draw epistemic conclusions from a general theory of linguistic representation. Any substantial account of meaning that treats it as determined by some aspect of our *use* will give a certain amount of epistemic privilege to that meaning-determining aspect, and such privileging may undermine various local or global skeptical worries. However, the antiskeptical advantages that use-based accounts of meaning bring often come at the cost of having to endorse antirealism about facts that we intuitively take to be more 'objective.'

For instance, 'descriptive' theories of meaning took what we meant by (many of) our terms to be determined by (some of) the descriptions we associate with them. 'Pen,' say, gets its meaning by having a description like 'cylindrical handheld artifact used to write with ink' associated with it. However, if such descriptions determined the meaning of the terms involved, then they could be known to be true in virtue of those meanings, and thus a purely 'linguistic' explanation of the possibility of *a priori* knowledge becomes available. We could know *a priori* that all pens are artifacts, because this proposition followed from the meaning of 'pen.' *A priori* truths were so knowable because they were analytic, that is, true in virtue of their meaning (Ayer, 1936).

In addition to the general doubts about analyticity stemming from Quine's work (1951), the description theory suffered from the fact that many of the propositions that it suggested that we knew *a priori* seemed, on reflection, not to be so knowable. There are, as Putnam stressed, things that we *actually* apply the term 'pen' to, and if *they* turned out not to be artifacts (say, they were alien organisms that were born rather than manufactured), then we would conclude that

pens were not artifacts, not that there were no pens (Putnam, 1975). The description theory showed how *a priori* knowledge would be possible, but doubts about such knowledge undermines the descriptive theory itself.

Descriptive theories made the mistake of tying what we mean exclusively to the general characterizations we associate with a term while ignoring what we have actually applied the term to (its 'putative extension'). This would suggest another approach to meaning, one that tied a term's meaning more closely to its putative extension. Such accounts can have antiskeptical consequences as well, but unlike descriptive theories, which seem better suited to underwrite *a priori* knowledge, such accounts underwrite perceptual knowledge.

Tying what a term means to its putative extension (or at least some privileged subset of it) is characteristic of the 'externalist' conceptions of meaning, which replaced the description theories in the 1970s (Kripke, 1972; Putnam, 1975). Such theories were often taken to rule out global skepticism about the external world, since if past usage determined what we meant by a term, then it made no sense to say that we might have always *misapplied* it. If, for instance, we were all brains in vats, and our 'perceptions' had always been determined by a supercomputer we were connected to, then our beliefs and assertions about the 'pens' around us would still be true, since by 'pen' we would mean the *virtual* pens that caused our pen utterances, not the 'real' pens that we hadn't had any contact with (Putnam, 1981). The externalist needn't rule out the kind of scenario that the skeptic traditionally relies on; she just denies that in such a scenario our beliefs would be false.

However, as an antidote to skepticism, such simple versions of externalism face two problems. First of all, even if externalism seems to secure the general veridicality of our beliefs, it does so at the expense of making our knowledge of our own thoughts problematic. After all, if the content of our beliefs is determined by such 'external' factors such as the nature of those items in our environment that we have *actually* applied it to in the past, and we don't have introspective access to such factors, then we may seem to lack introspective access to the content of our own beliefs (McKinsey, 1991; Falvey and Owens, 1994). We may have true beliefs, but our ignorance of their content prevents us from having knowledge. Second, it does seem possible that all of our applications of a term could be misapplications. For instance, there have been many witch sightings throughout history, and since there are no witches, we can assume that all these perceptually based witch beliefs were false. Our general characterization of 'witch' was not

satisfied by any of the people who were called 'witches.' The fact that the meanings of our terms could be affected by these general characterizations makes the transition from externalism to antiskepticism about perceptual judgments problematic. (The same problem emerged with the earlier 'paradigm case' arguments, which also made an attempt to equate a term's meaning with 'uncontroversial' cases of what we actually applied it to.)

The problems with these two sorts of account suggest that one's account of meaning should involve *both* a term's putative extension and the general characterizations associated with it. Something like this combination is found in Davidson's tying meaning to all of the sentences we take to be true, where this class includes both our general beliefs and our dispositions to apply terms to various objects in our actual environment. Davidson insists that on such a holistic account, we must treat a term's meaning as being that which would make *most* of the commitments associated with it true, so his view can still be understood as having antiskeptical consequences (Davidson, 2002). It might be that most of our beliefs about pens were false (because they weren't true of what we actually applied the term to), or it might be that most of the things we called 'pens' weren't pens (because they didn't satisfy most of our central 'pen beliefs'), but it couldn't be the case that *both* sorts of mistakes were the norm.

Still, even if most of our commitments turn out to be true on such an account, there will be no way to tell *a priori* which these will be. The holistic nature of the dynamic between our terms' putative extensions and general characterizations may insure that if our terms pick out anything at all, then the most deeply entrenched elements of our usage will turn out to be correct, but there is no guarantee that any *particular* aspect of a term's putative extension or general characterization will be among the set that ultimately reflects what we mean. Further, the fact that worries about our semantic self-knowledge remain on such an account, and that we can't be sure of the truth of any *particular* belief, suggests again that what the holistic theory gives us is a confidence that most of our beliefs must be true, not knowledge of any particular set of propositions.

It seems, then, that neither analyses of particular epistemic terms such as 'know' nor general semantic theories ruling out widespread misrepresentation can completely put skeptical worries to rest. This may be fortunate, since such strong antiskeptical consequences can be viewed as a sign that a semantic theory treats our representational capacities as *too* constrained and that a more expansive notion of what contributes to meaning is needed (Nagel, 1986).

See also: *A Priori* Knowledge: Linguistic Aspects; Analytic/Synthetic, Necessary/Contingent, and a *Priori/a Posteriori*: Distinction; Causal Theories of Reference and Meaning; Contextualism in Epistemology; Externalism about Content; Holism, Semantic and Epistemic; Meaning: Overview of Philosophical Theories; Ordinary Language Philosophy; Radical Interpretation, Translation and Interpretationalism; Reference: Philosophical Theories; Thought and Language: Philosophical Aspects; Truth Conditional Semantics and Meaning.

Bibliography

- Ayer A J (1936, 1946). *Language, truth and logic* (2nd edn.). New York: Dover.
- Burge T (1986). 'Cartesian error and the objectivity of perception.' In Pettit P & McDowell J (eds.) *Subject, thought, and context*. New York: Oxford University Press. 117–136.
- Burge T (1988). 'Individualism and self-knowledge.' *The Journal of Philosophy* 85, 649–663.
- Davidson D (2002). *Subjective, intersubjective, objective*. Oxford: Oxford University Press.
- DeRose K (1995). 'Solving the skeptical problem.' *The Philosophical Review* 104, 1–53.
- Falvey K & Owens J (1994). 'Externalism, self-knowledge, and skepticism.' *The Philosophical Review* 103, 107–137.
- Grice P (1989). *Studies in the way of words*. Cambridge: Harvard University Press.
- Jackman H (2001). 'Semantic pragmatism and a priori knowledge.' *Canadian Journal of Philosophy* 31, 455–480.
- Kripke S (1972, 1980). *Naming and necessity*. Cambridge: Harvard University Press.
- Malcolm N (1952). 'Knowledge and belief.' *Mind* 61(242), 178–189.
- McKinsey M (1991). 'Anti-individualism and privileged access.' *Analysis* 51, 9–16.
- Nagel T (1986). *The view from nowhere*. New York: Oxford University Press.
- Putnam H (1975). *Mind, language and reality*. New York: Cambridge University Press.
- Putnam H (1981). *Reason, truth and history*. New York: Cambridge University Press.
- Quine W V (1951). 'Two dogmas of empiricism.' In Quine W V (1953) *From a logical point of view*. Cambridge: Harvard University Press. 20–46.
- Sellars W (1956). 'Empiricism and the philosophy of mind.' In Sellars W (1963) *Science, perception and reality*. London: Routledge. 127–196.
- Strawson P F (1952). *Introduction to logical theory*. London: Methuen.
- Stroud B (1984). *The significance of philosophical skepticism*. Oxford: Oxford University Press.
- Williams M (1991). *Unnatural doubts*. Cambridge: Blackwell.
- Wittgenstein L (1969). *On certainty*. Oxford: Blackwell.

Essential Indexical

K Romdenh-Romluc, University of Nottingham, Nottingham, UK

© 2006 Elsevier Ltd. All rights reserved.

Indexicals are linguistic expressions such as 'I', 'here', and 'now' whose reference varies depending upon the context in which they are used. Many writers, including Castañeda (1966, 1967, 1968), Perry (1979), and Chisholm (1981), have argued that indexicals are irreducible to other modes of reference, such as names and definite descriptions. 'I' in my statement 'I am hungry,' e.g., cannot be replaced with the co-referring expression 'Komarine' without loss of cognitive impact, since I may believe that Komarine is hungry without realizing that Komarine is me. The same applies to uses of 'here' and 'now.' It is also claimed that the thoughts expressed using indexicals have a privileged role. The central insight is perhaps best appreciated if we consider Perry's claim that indexicals are essential for characterizing beliefs that motivate action (see also Castañeda, 1966, 1967, 1968).

I may see myself in a mirror and notice that the person I see reflected has dirt on her face, but unless I believe that *I* have a dirty face I will not wipe it. I may believe that the meeting starts at noon but will not be moved to act unless I believe that the meeting starts *now*. I may believe that the treasure is buried outside the old church, but I will not start digging unless I believe that the treasure is buried *here*. To act requires me to have beliefs about my surrounding environment and how I am situated with respect to it. Indexical beliefs provide for this because their content is essentially perspectival. It is this egocentric dimension that is lost if the indexical is replaced with a non-indexical term, which is why indexical expressions cannot be reduced to other modes of reference.

Perry (1979) famously argued that indexicals pose a problem for the traditional Fregean account of belief. According to this theory, beliefs are relations between subjects and propositions. The belief report 'Komarine believes that Fido is a dog' states that an individual – Komarine – stands in the relation of believing to the proposition 'Fido is a dog.' The

problem arises because the Fregean view implies (i) propositions are held to be true or false timelessly – not only true or false at a particular time, in a particular place, expressed by a particular person, and (ii) propositions are not individuated only by the ordinary objects that the proposition is about, but also by something like the conceptual components of the proposition that affect its cognitive impact. The propositions ‘Hesperus is Venus’ and ‘Phosphorus is Venus’ are about the same object, the planet Venus, but they are nevertheless different propositions since they are composed of different concepts. It is easy to see why indexicals pose a problem given the need to simultaneously satisfy (i) and (ii). A sentence such as ‘I am hungry’ does not pick out a proposition since its truth-value varies from context to context. To identify the proposition I believe when I utter ‘I am hungry,’ we need to specify some conceptual component associated with the term ‘I.’ To satisfy (i), the specification must yield a proposition that is timelessly true or false; to satisfy (ii), it must preserve the cognitive impact of ‘I am hungry.’ But as we saw above, there is no way of rendering ‘I am hungry’ as a proposition that is timelessly true or false without losing its cognitive impact. Various solutions to the problem have been proposed. Some attempt to refine the traditional account of belief (see, e.g., Evans, 1985), while others reject it in favor of a theory that can deal with the problem (see, e.g., Perry, 1986).

We sometimes attribute a use of an essential indexical to someone else. The usual way to do this is to use an *oratio obliqua* construction. I can, e.g., attribute the thought ‘I am going to be late for work’ to Sue by uttering, ‘Sue thinks that she herself is going to be late for work.’ Similarly, if Richard is in Malaga and thinks ‘It is hot here,’ I can attribute this thought to him by uttering ‘Richard is in Malaga; he says it is hot there.’ If Teresa thinks at 12 o’clock, ‘It is now time to go to the meeting,’ I can attribute this thought to her by uttering, ‘Teresa thought at 12 o’clock that it was then time to go to the meeting.’ (The alternative is to employ an *oratio recta* construction, e.g., ‘Sue thinks: I am going to be late for work.’) Castañeda (1966, 1967, 1968) labels

expressions that attribute a use of an essential indexical to someone else in *oratio obliqua* constructions ‘quasi-indicators.’ Like essential indexicals, quasi-indicators cannot be replaced with co-referring expressions without loss of cognitive impact. ‘Sue thinks that Sue is going to be late for work’ and ‘Sue thinks that she herself is going to be late for work’ attribute different thoughts to Sue, since she may have amnesia and forget that she is called ‘Sue.’ Thus, quasi-indicators raise the same kind of issues as essential indexicals. There are further problems associated with their use. Essential indexicals create difficulties for the traditional Fregean account of belief because the beliefs they express are essentially tied to the perspective of the believer. It is puzzling, therefore, how a listener can grasp the content of another’s indexical utterances, and further, convey that content to someone else. Again, there is disagreement concerning how the content of quasi-indexical attitudes – those expressed using quasi-indicators – should be understood.

Bibliography

- Castañeda H N (1966). ‘He’: a study in the logic of self-consciousness.’ *Ratio* 8, 130–157.
 Castañeda H N (1967). ‘Indicators and quasi-indicators.’ *American Philosophical Quarterly* 4, 85–100.
 Castañeda H N (1968). ‘On the logic of attributions of self-knowledge to others.’ *The Journal of Philosophy* 65, 439–456.
 Chisholm R (1981). *The first person*. Minnesota: University of Minnesota Press.
 Evans G (1985). ‘Understanding demonstratives.’ In Phillips A (ed.) *Collected papers*. New York: Oxford University Press. 291–321.
 Perry J (1979). ‘The problem of the essential indexical.’ *Nous* 13, 3–21.
 Perry J (1983). ‘Castañeda on ‘he’ and ‘I.’’ In Tomberlin J (ed.) *Agent language, and the structure of the world*. Indianapolis: Hackett Publishing Company. 15–39.
 Perry J (1986). ‘Thought without representation.’ *Supplementary Proceedings of the Aristotelian Society* 60, 137–152.

Event-Based Semantics

P Lasersohn, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

The notion of *events* may be used in semantic theory in a wide variety of ways, but the term *event-based semantics* normally refers to semantic analyses that

incorporate or adapt the proposal of Davidson (1967) that certain predicates take an implicit variable over events as an argument. In Davidson’s original proposal, this event argument is accommodated by analyzing the predicate as having one more argument place than is assumed in more traditional analyses. The event variable is existentially quantified, with the result that Sentence (1)a is assigned a

logical structure like (1)b, rather than the more traditional (1)c:

- (1) a. Jones buttered the toast.
- b. $\exists e$ butter(Jones, the toast, e)
- c. butter(Jones, the toast)

Thus, *butter* is analyzed as expressing a three-place relation between an individual who butters, an object that gets buttered, and a buttering event; and the sentence is analyzed as asserting that such an event exists.

The initial motivation for this proposal is that it provides a way to analyze adjuncts such as locative, temporal, and instrumental adverbial phrases. These are also treated as predicates of events – or more specifically as predicates of the same event as the verb. Each adjunct gives rise to its own clause in logical structure, and these are combined with the clause corresponding to the verb and its arguments by ordinary propositional conjunction. The existential quantifier binding the event variable takes scope over the whole structure, so that Sentence (2)a is assigned a logical structure like (2)b, for example:

- (2) a. Jones buttered the toast with a knife in the bathroom at midnight.
- b. $\exists e$ [butter(Jones, the toast, e) & with(e , a knife) & in(e , the bathroom) & at(e , midnight)]

This approach has an advantage over one in which the adverbials are treated as arguments of the verb, so that *butter* expresses a five-place relation as in (3):

- (3) butter(Jones, the toast, a knife, the bathroom, midnight)

If we adopt a formula like (3), but continue to represent *Jones buttered the toast* as in (1)c, with a two-place relation, we would seem to deny that *butter* expresses the same meaning in both sentences and claim instead that it is ambiguous. Nor is this a simple two-way ambiguity; *butter* will have to express different relations in each of the sentences in (4)

- (4) a. Jones buttered the toast with a knife.
- b. Jones buttered the toast in the bathroom.
- c. Jones buttered the toast at midnight.
- d. Jones buttered the toast with a knife at midnight.
- e. Jones buttered the toast with a knife in the bathroom.
- f. Jones buttered the toast at midnight in the bathroom.

This massive ambiguity is undesirable.

We might try to avoid the ambiguity by claiming that *butter* always expresses a five-place predicate, and that in examples in which fewer than five arguments appear overtly, the missing argument places are filled by implicit existentially bound variables.

However, this strategy ignores the fact that additional adverbials can always be added, with no limit on the number; as long as adverbials are analyzed as arguments, one cannot specify a fixed number of argument places for the verb, even if one allows for implicit arguments.

These problems are avoided completely under Davidson's proposal; *butter* is consistently analyzed as a three-place predicate. An unlimited number of adverbials may be added because these combine with the verb by ordinary conjunction, and not by filling argument places.

A second advantage to this analysis is that it correctly captures the fact that Sentence (2)a entails all the examples in (4) as well as (1)a, that (4)d entails (4)a, (4)c, and (1)a, and so on. Without some extra stipulation, these entailment relations do not fall out of an analysis in which adverbials are arguments to the verb. Extra stipulations are also required to capture these entailment relations in other alternative approaches to the semantics of adverbials, such as an approach in which they are treated as higher-order predicates taking verb intensions as arguments, as in (5):

- (5) [at-midnight(\wedge in-the-bathroom(\wedge with-a-knife(\wedge butter))))](Jones, the toast)

Davidson limited his original proposal to 'action sentences' and was explicit that it should not be applied to sentences such as $2 + 3 = 5$. Nonetheless, it is sometimes assumed that a similar analysis should be extended to some or all stative sentences (see especially Parsons, (1987/1988, 1990). In analyses employing both states and events, the term *eventuality* (introduced by Bach, 1986) is often used for the general category covering both.

The issue of which predicates have a hidden argument place for an eventuality, and which do not, is addressed in a well-known proposal by Kratzer (1995); see also Higginbotham (1983) and Fernald (2000). Kratzer suggests that *individual-level* predicates do not have such an argument place, and that *stage-level* predicates do. This position is supported by the following pattern of acceptability:

- (6) a. When Mary speaks French, she speaks it well.
- b. *When Mary knows French, she knows it well.
- c. When a Moroccan speaks French, she speaks it well.
- d. When a Moroccan knows French, she knows it well.

Assuming that indefinites contribute free variables to semantic representation (as in File Change Semantics or **Discourse Representation Theory**) and that *when*-clauses serve to restrict the domain of an implicit generic quantifier that can bind these variables,

the acceptability of (6)c–d is expected. The unacceptability of (6)b follows from a simple prohibition on vacuous quantification: The *when*-clause contains no free variables for the generic quantifier to bind. Why, then, is (6)a acceptable, as it does not contain any indefinite noun phrases either? Kratzer suggests it is because the stage-level predicate *speak* contributes a free Davidsonian event variable, whereas the individual-level predicate *know* does not.

Another area in which event variables have proven useful is in the semantics of perception reports (Higginbotham, 1983; Vlach, 1983; Parsons, 1990). Sentences like (7)a have been cited in support of thoroughgoing revisions to semantic theory of the kind adopted in **Situation Semantics**; but if we analyze this sentence as meaning that there is an event *e* of Mary's leaving, and an event *e'* of John's seeing *e*, we may assign it the logical structure in (7)b and obtain a reasonable analysis without using resources beyond those of ordinary first-order logic:

- (7) a. John sees Mary leave.
b. $\exists e[\text{leave}(\text{Mary}, e) \ \& \ \exists e' \text{ see}(\text{John}, e, e')]$

Davidson's technique of representing adjunct phrases as expressing separate logical clauses raises the possibility that major arguments of the verb such as the subject, direct object, and so on, might be treated in the same way. Davidson himself rejected this sort of extension, but a variant of it has been very popular in later work. Often termed the *Neo-Davidsonian* approach, this style of analysis treats the verb as a one-place predicate of eventualities; the subject, direct object, and so on do not serve directly as arguments to the verb, but instead stand in thematic relations to an event that fills the verb's sole argument place. A sentence such as (8)a thus receives a logical structure like that in (8)b:

- (8) a. Brutus stabbed Caesar.
b. $\exists e[\text{stab}(e) \ \& \ \text{agent}(e, \text{Brutus}) \ \& \ \text{patient}(e, \text{Caesar})]$

This approach to thematic relations appears first to have been proposed by Parsons (1980, 1985, 1990); see also Carlson (1984) and Krifka (1989, 1992). Note that this style of analysis requires an eventuality argument for all predicates that assign thematic roles, not just action predicates or stage-level predicates – at least on the assumption that thematic roles are represented in a uniform fashion for all predicates that assign them.

One advantage of this approach is that it allows a nice analysis of 'semantically optional' arguments: The direct object of *stab* may be omitted, as in (9)a; to give a logical form, we simply omit the clause for the corresponding thematic relation, as in (9)b:

- (9) a. Brutus stabbed.
b. $\exists e[\text{stab}(e) \ \& \ \text{agent}(e, \text{Brutus})]$

In a more conventional analysis, we might represent this sentence as in (10)a:

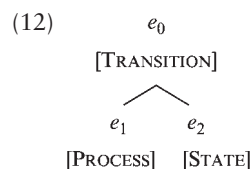
- (10) a. $\exists x \text{ stab}(\text{Brutus}, x)$
b. $\text{stab}(\text{Brutus})$

But as Parsons points out, this entails that Brutus stabbed something, whereas (9)a does not: Brutus could have stabbed and missed. If we try to avoid this entailment by representing (9)a as (10)b, we treat *stab* as ambiguous, expressing a different meaning in (9)a than it does in (8)a; but this is avoided in the Neo-Davidsonian analysis.

The idea that verbs are predicates of events has also been exploited in the analysis of certain types of nominalization (Higginbotham, 1985; Parsons, 1990). Combining the Neo-Davidsonian analysis with an assumption that nominals may express the same predicate of events as the verbs they derive from makes it possible to account for the validity of the argument in (11)a in a very straightforward fashion. This argument is represented as in (11)b, which is licensed by standard principles of first-order logic:

- (11) a. In every burning, oxygen is consumed.
Agatha burned the wood.
Therefore, oxygen was consumed.
b. $\forall e[\text{burn}(e) \rightarrow \exists e' [\text{consume}(e') \ \& \ \text{theme}(e', \text{oxygen}) \ \& \ \text{in}(e, e')]]$
 $\exists e[\text{burn}(e) \ \& \ \text{agent}(e, \text{Agatha}) \ \& \ \text{patient}(e, \text{the wood})]$
 $\exists e'[\text{consume}(e') \ \& \ \text{theme}(e', \text{oxygen})]$

Event arguments have also been used extensively in the analysis of Aktionsart. One line of research in this area, exemplified by Pustejovsky (1991, 1995) and Grimshaw (1990), represents events of certain complex types as structurally composed of events of simpler types. For example, an accomplishment predicate such as *build* may be associated with events of the structure illustrated in (12):



Here *e*₁ represents the building process itself, whereas *e*₂ represents the resultant state of something having been built. As a telic predicate, *build* involves reference not just to the building process, but also to its culmination in the transition to a result state, represented by *e*₀.

A rather different approach to the event-theoretic representation of Aktionsart was developed by Krifka

(1989, 1992). Here, a 'sum' operation is assumed on events, so that for any two events e_1, e_2 , a complex event $e_1 \sqcup e_2$ consisting of them is assumed to exist. A part/whole relation is definable in terms of the sum operation: $e_1 \sqsubseteq e_2$ (' e_1 is a part of e_2 ') iff $e_2 = e_1 \sqcup e_2$. Predicates are assumed to denote sets of events, allowing aspectual classes to be defined in terms of closure conditions on these sets. Cumulative predicates are those denoting sets that are closed under the sum operation:

$$(13) \text{ CUM}(P) \leftrightarrow \forall x, y [P(x) \ \& \ P(y)] \rightarrow P(x \sqcup y)]$$

For example, if x is a walking event, and y is a walking event, their combination is also a walking event. In contrast, quantized predicates denote sets from which proper parts of their members are excluded:

$$(14) \text{ QUA}(P) \leftrightarrow \forall x, y [P(x) \ \& \ P(y)] \rightarrow x \not\sqsubseteq y]$$

For example, if x is an event of drinking a glass of wine, and y is also an event of drinking a glass of wine, x cannot be a proper part of y . Cumulative and quantized predicates of events correspond roughly to the familiar categories of atelic and telic predicates, respectively. However, by assuming a sum operation and corresponding part/whole relation on individuals, and not just events, it is possible to apply these concepts to predicates of individuals as well. For example, if x is wine and y is wine, their sum must also be wine, establishing *wine* as a cumulative predicate; if x is a glass of wine and y is a glass of wine, x may not be a proper part of y , establishing *glass of wine* as a quantized predicate. The status of a verb's arguments as cumulative or quantized often affects the aspectual category of the verb phrase or sentence; hence (15)a is cumulative whereas (15)b is quantized:

- (15) a. John drank wine.
b. John drank a glass of wine.

Assuming a Neo-Davidsonian representation of thematic relations, Krifka explored the mathematical properties such relations must have to give this effect.

A related application of part/whole structures for events was developed by Lasnik (1990, 1995) for the semantics of plurality (see also Schein, 1993). Distributive readings of predicates with plural subjects are analyzed as representing events that divide into smaller events corresponding to the individual members of the group denoted by the subject. For example, an event of John and Mary sitting divides into a smaller event of John sitting and one of Mary sitting. Collective readings correspond to events that cannot be divided in this way; an event of John and Mary being a happy couple does not divide into an

event of John being a happy couple and an event of Mary being a happy couple. Representing the collective/distributive distinction in terms of event structure in this way makes possible an extensional analysis of adverbials like *together*, which imposes a collective reading on the predicates it modifies.

Event-based semantics has also been fruitfully applied to a wide variety of other problems, of which space limitations prevent a discussion here: plurality (Lasnik, 1990, 1995; Schein, 1993; Landman, 2000), temporal anaphora and narrative progression (Hinrichs, 1986; Partee, 1984), cognate objects (Mittwoch, 1998), adjectives (Larson, 1998), and many others. There is also a large philosophical literature on events, much of which relates directly to Davidsonian-style event-based semantics; see Davidson (1980), LePore and McLaughlin (1985), Higginbotham *et al.* (2000), and the references cited therein.

See also: Discourse Representation Theory; Plurality; Situation Semantics.

Bibliography

- Bach E (1986). 'The algebra of events.' *Linguistics and Philosophy* 9, 5–16.
- Carlson G (1984). 'Thematic roles and their role in semantic interpretation.' *Linguistics* 22, 259–279.
- Davidson D (1967). 'The logical form of action sentences.' In Rescher N (ed.) *The logic of decision and action*. Pittsburgh: University of Pittsburgh Press. 81–94. Reprinted in Davidson (1980), 105–122.
- Davidson D (1980). *Essays on actions and events*. Oxford: Oxford University Press.
- Fernald T (2000). *Predicates and temporal arguments*. Oxford: Oxford University Press.
- Grimshaw J (1990). *Argument structure*. Cambridge, MA: MIT Press.
- Higginbotham J (1983). 'The logic of perceptual reports: an extensional alternative to Situation Semantics.' *Journal of Philosophy* 80, 100–127.
- Higginbotham J (1985). 'On semantics.' *Linguistic Inquiry* 16, 547–593.
- Higginbotham J, Pianesi F & Varzi A C (2000). *Speaking of events*. Oxford: Oxford University Press.
- Hinrichs E (1986). 'Temporal anaphora in discourses of English.' *Linguistics and Philosophy* 9, 63–82.
- Kratzer A (1995). 'Stage-level and individual-level predicates.' In Carlson G N & Pelletier F J (eds.) *The generic book*. Chicago: University of Chicago Press. 125–175.
- Krifka M (1989). 'Nominal reference, temporal constitution, and quantification in event semantics.' In Bartsch R, van Benthem J & van Emde Boas P (eds.) *Semantics and contextual expression*. Dordrecht: Foris Publications. 75–115.

- Krifka M (1992). 'Thematic relations as links between nominal reference and temporal constitution.' In Sag I A & Szabolcsi A (eds.) *Lexical matters*. Stanford, CA: Center for the Study of Language and Information. 29–53.
- Landman F (2000). *Events and plurality: the Jerusalem lectures*. Dordrecht: Kluwer Academic Press.
- Larson R (1998). 'Events and modification in nominals.' In Strolovitch D & Lawson A (eds.) *Proceedings from semantics and linguistic theory VIII*. Ithaca, NY: CLC Publications. 145–168.
- Lasersohn P (1990). 'Group action and spatio-temporal proximity.' *Linguistics and Philosophy* 13, 179–206.
- Lasersohn P (1995). *Plurality, conjunction and events*. Dordrecht: Kluwer Academic Publishers.
- LePore E & McLaughlin B (eds.) (1985). *Actions and events: perspectives on the philosophy of Donald Davidson*. Oxford: Basil Blackwell.
- Mittwoch A (1998). 'Cognate objects as reflections of Davidsonian event arguments.' In Rothstein S (ed.) *Events and grammar*. Dordrecht: Kluwer Academic Publishers. 309–348.
- Parsons T (1980). 'Modifiers and quantifiers in natural language.' *Canadian Journal of Philosophy* 6(suppl.), 29–60.
- Parsons T (1985). 'Underlying events in the analysis of English.' In LePore E & McLaughlin B (eds.) *Actions and events: perspectives on the philosophy of Donald Davidson*. Oxford: Basil Blackwell. 235–267.
- Parsons T (1987/1988). 'Underlying states in the semantic analysis of English.' *Proceedings of the Aristotelian Society* 88, 13–30.
- Parsons T (1990). *Events in the semantics of English: a study in subatomic semantics*. Cambridge, MA: MIT Press.
- Partee B (1984). 'Nominal and temporal anaphora.' *Linguistics and Philosophy* 7, 243–286.
- Pustejovsky J (1991). 'The syntax of event structure.' *Cognition* 41, 47–81.
- Pustejovsky J (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Rothstein S (ed.) (1998). *Events and grammar*. Dordrecht: Kluwer Academic Publishers.
- Schein B (1993). *Plurals and events*. Cambridge, MA: MIT Press.
- Tenny C & Pustejovsky J (eds.) (2000). *Events as grammatical objects*. Stanford, CA: CSLI Publications.
- Vlach F (1983). 'On situation semantics for perception.' *Synthese* 54, 129–152.

Evolution of Semantics

V Evans, University of Sussex, Brighton, UK

© 2006 Elsevier Ltd. All rights reserved.

One of the most important functions of language is to facilitate the 'transmission' of thought from one language user to another. A number of scholars, including Sperber and Wilson (1995), and Tomasello (1999, 2003), have observed that verbal communication requires both a code – which is to say a language-system involving conventional symbols, pairings of form and meaning – and intentional mechanisms such as inference-reading abilities. While both these aspects are essential for verbal communication, communication can, in principle, occur in the absence of a code. Indeed, as we shall see, intentionality and the ability to recognize communicative intentions are likely to have been necessary prerequisites for the evolution of symbolic representation in language.

To function as a means of communication, an important prerequisite of a code, which is to say a language-system, is to be able to encode and externalize humanly-relevant concepts and combinations of concepts. Semantic knowledge, therefore, concerns the range and nature of humanly relevant concepts that can be expressed in language, and the way language serves to combine concepts in order to convey

complex ideas. In this article, we explore (i) possible cognitive preadaptations for the development of semantic knowledge, and (ii) the range and nature of conceptual structure as encoded in language, and suggestions as to the way that this structure may have evolved.

Unlike some other aspects of language, there is scant evidence we can draw on in attempting to reconstruct the evolution of semantic knowledge. After all, we are, in essence, attempting to reconstruct the evolution of human cognition. To do this, we are relying on indirect evidence drawn from primatology and comparative psychology, paleontology, evolutionary anthropology, and evolutionary psychology. Nevertheless, in view of some recent developments in linguistics, both in terms of uncovering and better understanding semantic phenomena, and recent theory-construction, we can now construct some plausible paths of semantic evolution that will at least facilitate further inquiry.

Cognitive Preadaptations for Semantic Knowledge

Language is characterized by being representational or 'symbolic.' That is, a language consists of a structured set of 'symbolic units' consisting of form and

meaning components. While this definition represents the received view for lexical items, a growing body of scholarship argues that grammatical patterns can also be thought of as being inherently symbolic in nature (Langacker, 1987). Symbolic units consist of two further units: a phonological unit and a semantic or conceptual unit. The semantic unit, which is what we are concerned with here, has been variously termed a 'lemma' (Levelt, 1989) or a 'lexical concept' (Evans, 2004). In this section, we approach the evolution of semantic knowledge in a general way by considering the cognitive preadaptations that may have paved the way for the emergence of semantic knowledge.

The Importance of Motor Evolution

Donald (1991, 1999) has argued that there were two essential prerequisites for the evolution of symbolic units. One defining characteristic of language is that it can represent a particular idea or entity in the absence of a concrete cue: the design feature of language known as 'displacement.' For this representation to occur, hominids had to gain conscious access to their own memories (Donald, 1999). A second and crucial preadaptation for the emergence of language was the development of voluntary motor control. That is, hominids must have developed the ability to attend to their own action patterns, and to select, trigger, and 'edit' action pattern sequences. According to Donald, this development gave rise to 'mimesis,' a form of nonlinguistic representation. Mimetic action is representational in that it relies on perceptual resemblance to represent itself. For instance, hominid tool use, which can be traced back 1.5 million years, may have employed mimetic representation not only for showing and learning how to employ a tool, but through 'editing' motor routines through rehearsal, to improve the way in which the tool was used. Forms of representation such as mime, dance, ritual acts, and some kinds of music are also mimetic, serving as a form of communication that is nonlinguistic in nature. According to Donald, mimetic action was the earliest form of communication, upon which the later development of language may have been built.

While voluntary control of the musculature must have been important in the rise of this early and basic form of communication, and presumably also facilitated the later development of phonetic abilities and phonological systems, for Donald, linguistic representation is of a different kind from mimetic representation. While mimetic representation is holistic, a key characteristic of semantic knowledge, as represented by the inventory of lexical concepts available in the languages of the world, is that symbolic units

serve to 'parse' sensory or perceptual experience into component parts, e.g., tree versus rock versus mountain, and even to encode a particular perspective with respect to which a component is viewed. For, instance, 'shore' and 'coast' both encode the same strip of land at the edge of the sea, but do so from different perspectives. Thus, for Donald, the importance of mimetic representation was that it created an appropriate cultural context, what he terms 'mimetic culture,' in which communication took place, and more precise disambiguation could occur with the advent of linguistic representation.

The Importance of Intention-Reading Skills

Another important preadaptation for the development of semantic knowledge is likely to have been the emergence of the ability to read intentions. According to Tomasello (1999), this sort of ability was the crucial preadaptation required for the evolution of symbolic abilities such as language more generally. Research in developmental psychology reveals that during early ontogeny, shortly before a year old, human infants begin to experience themselves as 'intentional agents.' That is, they perceive themselves as beings whose attentional and behavioral strategies are goal-directed. Accordingly, human infants also come to see others with whom they identify, conspecifics, as intentional agents. Crucially, it is shortly after this ontogenetic 'breakthrough' that language begins to emerge (Tomasello, 2003). Later, from around the age of three, human infants begin to develop the notion of themselves and conspecifics as 'mental agents.' This development constitutes the emergence of the 'theory-of-mind,' in which children develop the ability to conceive that others can hold different views from their own.

The importance of viewing oneself and conspecifics as intentional agents is far-reaching. From this view, it follows that others are intentional agents who possess mental states that can be directly influenced and manipulated. For instance, pointing at an object can cause one intentional agent – who recognizes the person doing the pointing as an intentional agent attempting to direct attention – to follow the direction of pointing and thus share a 'joint attentional frame' (Tomasello, 1999, 2003). Thus, from this perspective, the importance of a lexical concept being associated with a particular linguistic form is in the utility of the symbolic unit in affecting the mental state of another in some way, such as by coordinating behavior. In other words, language, and the lexical concepts encoded by language, require intention-reading skills, which derive from the awareness that conspecifics represent intentional agents whose mental states can be influenced and manipulated by language.

A number of scholars view intention-reading abilities as an outcome of earlier evolutionary developments. For instance, Whiten (1999) argued that intention-reading skills constitute the outcome of the emergence of what he termed 'deep social mind.' This result can be characterized by cooperative behaviors including the sharing of food, monogamous reproduction – which has been claimed to be the ancestral pattern for humans – and behavior such as communal hunting. Indeed, Whiten argued that intention-reading abilities would have been essential for coordinating activities such as hunting, success at which requires being able to read the intentions of cohunters, and possibly also the prey.

Intention-reading skills most likely evolved by reading observables, such as direction of gaze, direction of motion, and so on. Thus, intention-reading skills are likely to have emerged from behavior-reading skills. On some accounts, chimpanzees are capable of rudimentary intention-reading abilities. Thus, intention-reading might be more than 6 million years old (Byrne, 1999), the time when hominids and chimpanzees separated.

Some scholars have argued that intention-reading in hominids can be viewed as a consequence of a long chain of evolutionary development. For instance, Savage-Rumbaugh (1994) suggested that bipedalism may have set in chain a series of evolutionary developments that gave rise to the cognitive ability to take the perspective of others (intention-reading). For instance, while chimpanzees and gorillas are distinguished from orangutans by a kind of quadrupedal locomotion termed 'knuckle-walking,' early hominids, the australopithecines, who emerged sometime between 4 and 5 million years ago, were distinguished by bipedalism. According to Savage-Rumbaugh, knuckle-walking and bipedalism were distinct and independent solutions to traversing open terrain and transporting infants. However, a consequence of bipedalism, but not knuckle-walking, is that the parent would have had to pay more attention to the infant, which is carried in the arms. In particular, the parent must remember to pick the child up after it has been put down. This consequence may have led to the later evolution of being able to take the perspective of others.

Similarly, Byrne (1999) argued that there may be more remote evolutionary antecedents for intention-reading abilities. One hypothesis is that our relatively large brachiating ancestors, for whom a fall would have been deadly, may have accomplished arboreal locomotion by advance planning. The mental representation of self as an entity moving through space would have prefigured representational abilities in general, and would have facilitated planning

Table 1 Human intention reading abilities

Human intention reading abilities include . . .

The ability to coordinate or share attention, as when an infant and adult both attend to the same object

The ability to follow attention and gesturing, as when an infant follows an adult's pointing or gaze, in order to attend to an object

The ability to actively direct attention of others, such as drawing attention to a particular object or event, for instance, through pointing

The ability of culturally (imitatively) learning the intentional actions of others, such as imitating verbal cues in order to perform intentional actions such as declarative, interrogative or imperative speech functions

a trajectory of motion. Self-representation, and the ability to consciously plan one's movements are cognitive achievements that imply intentionality, and the later evolution of intention-reading skills. The suite of intention-reading skills evident in modern humans is summarized in Table 1.

The Importance of Personality Types

This issue concerns the idea that the earliest lexical concepts may have related to personality traits (King *et al.*, 1999). Recent research suggests that personality traits are stable across time and between contexts, correlate with verbal and nonverbal behaviors, and can be reliably judged by human observers. Moreover, King *et al.* (1999) argued that such behaviorally-signaled personality traits as reliability, dominance, and trustworthiness are directly relevant to complex social interactions involving competition, cooperation, sharing, sexual selection, and so on. King *et al.* (1999) suggested that it is the context-independent nature of such complex personality traits, and their importance for hominids that suggests such traits may have been encoded as the earliest lexical concepts.

For instance, studies that have sought to teach chimpanzees to manipulate symbolic units have found that for symbol use to succeed, meaning must be de-contextualized. Consider the example of an apple. If a symbol is applied to this referent, it is not clear which properties of the scene the symbolic form relates to. For instance, it could refer to the apple's color, shape, or that it is an item of food. Until the referent has been experienced in a number of contexts, it is not clear which aspect of the referent is being indexed, and thus what the lexical concept is that is being associated with the form. As personality traits are context-independent and readily identifiable by observers, then an early linguistic form that indexed a particular personality trait might have served as an early lexical concept. That is, personality traits achieve the displacement aspect of lexical concepts by virtue of being inherently context-independent.

For this reason, symbolic representation in language may have taken personality traits as the first lexical concepts.

The Nature and Evolution of Semantic Knowledge

In this section, we examine the nature of semantic knowledge in more detail. That is, we examine how humans organize the world and their experience of the world into concepts. We also speculate on possible evolutionary bases of semantic knowledge of this kind and the cognitive mechanisms underlying this knowledge.

Concept Formation

‘Semantic structure’ constitutes the meaning system directly expressed by and encoded in language. In other words, semantic structure is the form that conceptual structure takes for expression in language. Thus, in order to get a sense of the nature of semantic knowledge, for instance, the nature and range of lexical concepts, we must begin by examining the nature of conceptual structure. In this section, then, we consider the basic units of concept structure, ‘concepts.’ We consider the following question: Where do concepts come from?

For psychologists, concepts are the basic units of knowledge and are essential both for ‘categorization’ – the ability to identify individuals, entities, and instances – and ‘conceptualization’ – the ability to construct alternative perspectives (Barsalou, 1992). To illustrate the notion of conceptualization, consider the sentences in (1) and (2). Each provides a different conceptualization of the concept Book:

- (1) That book is heavy.
- (2) That book is boring.

While the example in (1) relates to the book ‘as tome,’ the example in (2) relates to book ‘as text.’

Since the work of the French philosopher René Descartes in the 17th century, who developed the principle of Mind/Body dualism, there has been a common assumption within philosophy and, more recently, the other cognitive sciences, that conceptual structure can be studied without recourse to the body, and hence without recourse to ‘embodiment.’ In modern linguistics, this ‘objectivist approach’ has been most evident in the approach to meaning known as ‘Formal Semantics.’ Proponents of this approach assume that it is possible to study meaning as a formal or computational system without taking into account the nature of human bodies or human experience. This position is problematic from an evolutionary

perspective as it entails that a new discontinuous cognitive adaptation was required for conceptual structure. Conceptual structure, on this account, is assumed to employ what has been termed an ‘amodal’ (nonperceptual) form of representation. Amodal representation is distinct from the ‘modal’ or perceptual forms of representation that presumably had to exist prior to the emergence of conceptual structure, in order to represent ‘percepts’ (Barsalou, 1999).

The last two decades or so have seen a shift from modeling conceptual representation in terms of amodal systems, towards a more perceptual-based or ‘embodied perspective.’ An embodied perspective takes the view that concepts derive from percepts, and thus, conceptual structure is fundamentally perceptual in nature. Within linguistics, this general perspective has been advocated most notably by Lakoff and Johnson (1980, 1999; Lakoff, 1987), and also by Jackendoff (1983, 1992, 2002).

In general terms, the idea is that concepts have an embodied character. This idea constitutes the thesis of embodied cognition (see Ziemke, 2003 for discussion).

The idea that concepts are embodied assumes that we have a species-specific view of the world, due to the nature of our physical bodies.

One obvious way in which our embodiment affects the nature of experience is in the realm of color. While the human visual system has three kinds of photoreceptors or color channels, other organisms often have a different number. For instance, the visual system of squirrels, rabbits, and possibly cats, makes use of two color channels, while other organisms, for instance, goldfish and pigeons, have four color channels (Varela *et al.*, 1991). Having a different range of color channels radically alters how the world of color is perceived. This difference affects our experience of color in terms of the range of colors accessible to us along the color spectrum. Moreover, while some organisms can see in the infrared range, humans are unable to see in this range (Jackendoff, 1992). It’s clear, then, that the nature of our visual apparatus – an aspect of our physical embodiment – determines the nature and range of our visual experience. The position that different organisms have different kinds of experiences due to the nature of their embodiment is known as ‘variable embodiment.’

The position that our experience is embodied – that is, structured in part by the nature of the kinds of bodies/neuro-anatomical structure we have – has consequences for conceptual structure. This corollary follows because the concepts we have access to, and the nature of the ‘reality’ we think and talk about, is a function of our embodiment. In other words, we can only talk about what we can perceive and

think about, and the things that we can perceive and think about derive from embodied experience. Hence, the human mind must bear the imprint of embodied experience.

Some psychologists have made specific proposals as to how embodied experience gives rise to concepts. For instance, the developmental psychologist Jean Mandler (2004) suggested that through a process of ‘perceptual meaning analysis,’ percepts come to be re-coded as concepts. Mandler argued that this process occurs alongside percept formation and begins virtually from birth. However, she viewed percepts and concepts as wholly distinct forms of representation. Another view has been proposed by Barsalou (1999). He argued that a concept is akin to a remembered perceptual state, which he termed a ‘perceptual symbol.’

From an evolutionary perspective, if it is correct that concepts are fundamentally perceptual in nature, then by virtue of early hominids gaining conscious access to the contents of their own memories, little additional complexity in terms of cognitive development is required for a rudimentary conceptual system to have emerged. This corollary follows as concepts, on this account, are something akin to ‘remembered percepts.’

The Nature of Lexical Concepts: The Natural Partitions Hypothesis

Having examined conceptual structure, we now turn to semantic structure.

Linguists have traditionally classified lexical concepts into those that are encoded by ‘open’ versus ‘closed class forms.’ Open class forms include, for English, nouns, verbs and adjectives, while closed class forms include determiners, prepositions, conjunctions, and so on. The basic insight is that it is much harder to add new members to the closed class set than to the open class set. A related insight is that open class forms tend to have much richer denotational meaning, while closed class forms are associated with lexical concepts that have more schematic or relational meaning. That is, they provide connections to other lexical concepts that have a more referential meaning.

However, since at least the early 1980s, the strict separation between closed and open class concepts

has been called into question. This query stems from the observation that the division between open and closed class concepts constitutes more of a continuum rather than a strict bifurcation. For instance, Gentner (1981) pointed out that verbs, which are normally thought of as being open class, are highly relational in nature, a feature associated with closed class elements.

More recently, Gentner and Boroditsky (2001) have elaborated on this view, suggesting that open class lexical concepts exhibit ‘cognitive dominance.’ This contrasts with closed class concepts that exhibit ‘linguistic dominance.’ These notions relate to the similar idea expressed by Langacker (1987), who used the terms ‘conceptually autonomous’ versus ‘conceptually dependent.’ The basic idea is that lexical concepts associated with prototypical open class (autonomous) forms obtain their reference independently of language, which is to say from the world, while prototypical lexical concepts associated with closed class or relational forms obtain their reference from language. Moreover, whether a form is cognitively dominant (or autonomous) or linguistically dominant (or dependent) is a matter of degree. A proposed continuum is given in Figure 1.

In order to account for the cognitive dominance of prototypical open class lexical concepts (i.e., nouns), Gentner (1981) proposed the Natural Partitions Hypothesis. This idea holds that concepts that are encoded as prototypical open class elements such as individuals and objects are ‘individuated.’ That is, entities of this kind constitute densely bundled collections of percepts. Thus, an entity such as a rock or a tree ‘stands out.’ In Gestalt Psychology terms, a rock constitutes the figure in the figure-ground organization of a given scene. The Natural Partitions Hypothesis states that certain aspects of the world are given by the world. These entities are typically encoded crosslinguistically by nouns, and are acquired first by children. On this account then, bundles of percepts are ‘given’ by the world, and are simply labeled by language.

The Natural Partitions Hypothesis offers an intriguing insight into a possible order of evolution among lexical concepts – which is to say concepts encoded by language. That is, we might speculate, based on this, that the very first lexical concepts were those for

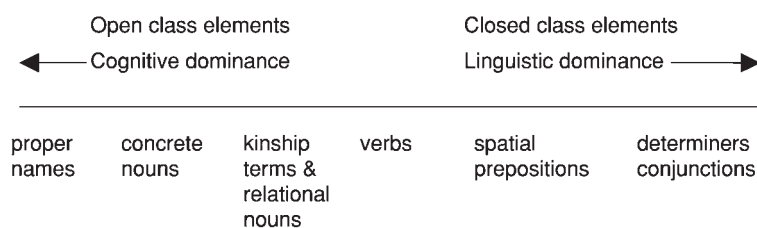


Figure 1 Division of dominance among form classes of lexical concepts. (Adapted from Gentner and Boroditsky, 2001: 216.)

individuals, including animals (and possibly classes of animals) and objects. Concepts of this kind have the most cognitive dominance. That is, they have highest conceptual autonomy. Other lexical concepts may have evolved later.

Further, there is a correlation between the position of a lexical concept on the continuum of dominance (see Figure 1) and the form class associated with the lexical concept. Although this correlation is not exact, for instance, ‘destruction’ and ‘destroy’ encode a similar concept employing different lexical classes (noun versus verb), it is plausible that the evolution of lexical classes (or ‘parts of speech’) emerged due to distinctions in the relative dominance or autonomy being further, later, encoded by morphosyntactic properties of language.

Lexical Concepts and Concept-Combination

From an evolutionary perspective, being able to form concepts and express them via language, while a remarkable achievement, doesn’t begin to approach the range and complexity of the semantic structure available to modern Homo Sapiens. Lexical concepts are only a subset of our semantic knowledge. Another important aspect of semantic knowledge concerns our ability to combine lexical concepts in order to give rise to new and different kinds of conceptual structure. Moreover, it is a striking fact that concept combination produces complex concepts that are not simply the sum of the individual parts that comprise the derived concept. For instance, the complex concept *Petfish* is not simply the intersection of the concepts *Pet* and *Fish*. Rather, the concept *Petfish* has its own concept-internal structure, known as ‘category structure.’

For instance, while most people would rank mackerel, which is silver in color, as a good example of the *Fish* category, a cat or a dog would be rated as a good example of the *Pet* category. Yet, a good example of a *Petfish* is a goldfish. Not only is a goldfish not silver, it is not soft and cuddly either. An important task in developing an evolutionary perspective on semantic knowledge is to account not only for the way in which lexical concepts are formed, but also for the mechanisms responsible for concept combination.

A recent approach to concept combination of this kind argued that complex concepts result from a process of ‘conceptual integration’ (Fauconnier and Turner, 2002; Turner and Fauconnier, 1995). This process involves what is termed ‘selective projection’ of content from each of the concepts that give rise to the complex concept, as well as additional material derived from background knowledge, such as knowledge that the kinds of fish we keep in fishbowls are typically goldfish. This process is termed ‘completion.’ Thus, the complex concept, known as a ‘conceptual blend,’ has structure associated with it that is found in neither of the ‘input’ concepts that give rise to it. This structure is diagrammed in Figure 2.

Clearly, some form of conceptual integration allows humans to combine and manipulate concepts in order to produce more complex ideas. Fauconnier and Turner argued that the emergence of cognitively modern human beings, during the upper paleolithic era, somewhere in the region of 50 000 years ago, points to the development of a new cognitive ability: our ability to perform conceptual integration. While anatomically modern humans appear to have existed from at least 100 000 years ago, the upper paleolithic stands out. This period witnessed the emergence of

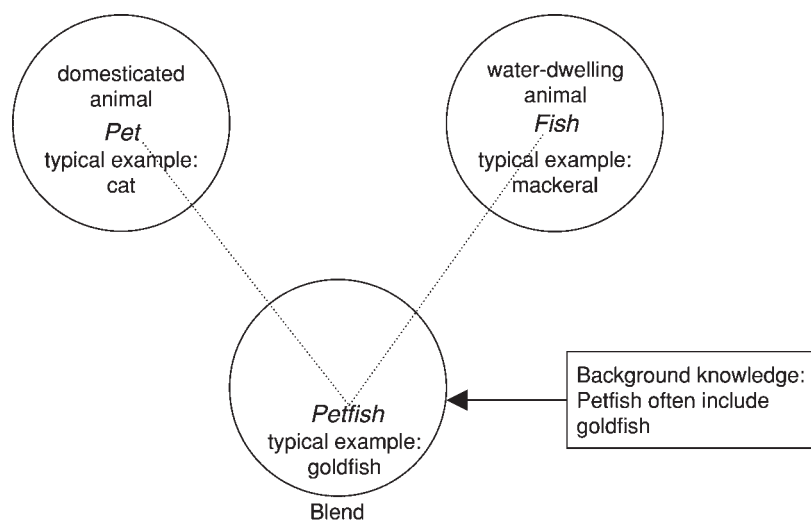


Figure 2 Conceptual integration for the composite concept *goldfish*.

new social and technological breakthroughs, including the development of projectile points made from bony material for use in hunting, the manufacture of personal adornments, the development of sophisticated art, evidence of belief systems such as religion and magic, plus manmade shelters were built for the first time, sewn clothing was worn, and sculptures were produced. Fauconnier and Turner argued that what made advances such as these possible, was that humans had evolved the ability to perform complex conceptual integrations. This process, then, may have facilitated composing and elaborating concepts to produce new and more elaborate conceptual structures.

Polysemy

Another striking aspect of semantic knowledge is the phenomenon of ‘polysemy.’ This aspect constitutes the way in which a range of related lexical concepts can be expressed using a single form. For instance, the English preposition ‘over’ has a number of distinct but related lexical concepts associated with it. Consider some of the distinct lexical concepts proposed by Tyler and Evans (2003):

- (3a) The picture is over the sofa [‘above’]
- (3b) The picture is over the hole [‘covering’]
- (3c) The ball is over the wall [‘on-the-other-side-of’]
- (3d) She has a strange power over him [‘control’]
- (3e) The government handed over power [‘transfer’]
- (3f) She prefers wine over beer [‘preference’]
- (3g) The relationship is over [‘completion’]
- (3h) The relationship evolved over the years [‘temporal’]
- (3i) The fence fell over [‘reflexive’]
- (3j) They started the race over [‘repetition’]

Recent research has argued that polysemy, far from being merely a ‘surface’ phenomenon, is in fact conceptually real. That is, polysemy patterns reflect distinct lexical concepts, stored as different senses in the mental lexicon (Evans, 2004; Lakoff, 1987; Tyler and Evans, 2003). Accordingly, from an evolutionary perspective, the challenge is to explain how the proliferation of lexical concepts, i.e., polysemy, arises.

A recent perspective is that polysemy emerges from the interaction between language use and contexts of use, due to the conventionalization of situated (or invited) inferences (Traugott and Dasher, 2002; Tyler and Evans, 2003; Evans, 2004). For instance, the ‘covering’ meaning associated with ‘over’ may have derived from contexts of use in which, in a given spatial scene, an element placed above another entity thereby covered it. Through a process of decontextualization, the ‘covering’ meaning was reanalyzed as

being a distinct meaning component. Once this reanalysis occurred, it could be used in novel ways unsupported by the original spatial scene that gave rise to the inference in the first place (Tyler and Evans, 2003).

From an evolutionary perspective, the importance of polysemy and meaning-extension is that it illustrates how language, in conjunction with human experience, can give rise to new lexical concepts. Moreover, this particular phenomenon of meaning-extension illustrates how language can flexibly increase its repertoire of lexical concepts without increasing the number of linguistic forms.

Abstract Concepts

Another important aspect of semantic structure relates to so-called abstract concepts. These include lexical concepts such as Truth, Justice, or Theory. Concepts of these kinds are abstract in the sense that they cannot be straightforwardly accounted for in terms of perceptual recording, precisely because it’s not clear what their perceptual basis is, and even whether they have one. Indeed, abstract concepts provide a significant challenge if we are to attempt to provide an evolutionary account maintaining the thesis of embodied cognition.

An influential framework that provides an account that is based in perceptual or embodied experience is the ‘conceptual metaphor theory’ of Lakoff and Johnson (1980, 1999). Lakoff and Johnson argued that abstract concepts are grounded in embodied experience, and thus our perception of the world, even if the grounding is not direct. This grounding is achieved by virtue of ‘conceptual metaphors,’ which are long-term conceptual mappings that serve to project structure from a ‘source concept,’ which relates to perceptual experience onto the abstract concept, the ‘target concept.’ For instance, we commonly understand the abstract concept of Quantity in terms of the more perceptually concrete concept of Verticality, as evidenced by examples such as the following:

- (4a) The price of stocks has **gone up**.
- (4b) Her score is **higher** than mine.

In both these examples, an abstract notion of Quantity is understood in terms of physical position or motion on the vertical axis. This understanding is licensed by the conceptual metaphor Quantity Is Vertical Elevation.

The most recent version of conceptual metaphor theory recognizes two distinct kinds of conceptual metaphors: ‘primary metaphors,’ which are directly grounded in experience and constitute ‘primitive’ conceptual mappings, and more complex ‘compound metaphors,’ which are constructed out of the more

experientially basic primary metaphors (Grady, 1997; Lakoff and Johnson, 1999).

For instance, when we understand Theories in terms of Physical Structures, as evidenced by the following examples:

(5a) Is that the **foundation** for your theory?

(5b) The argument is **shaky**.

Grady argues that the motivation for linguistic examples such as these is in fact two primary metaphors, Persisting Is Remaining Erect and Organization Is Physical Structure. These unify to give the compound metaphor An Abstract Organized Entity [such as a theory] Is An Erect Physical Object (Grady, 1997).

Thus, it is only primary metaphors that are grounded in perceptual experience. The motivation for the conceptual associations captured by primary metaphors is due to a tight and ubiquitous correlation in experience. For instance, there is a tight and recurring correlation in experience between quantity and height. When we fill a glass with water, an increase in quantity correlates with an increase in height. Thus, primary metaphors are motivated by 'experiential correlation.'

From an evolutionary perspective, the phenomenon of 'metaphoric' mappings holding between concepts from different parts of 'conceptual space,' known as 'domains,' allows us to account for how perceptual information can be recruited in order to construct more abstract concepts, such as Quantity and Theories. This phenomenon suggests that, in addition to being able to recode percepts as concepts and combine concepts, the conceptual system must have additionally developed a mechanism for projecting structure from one conceptual domain to another in order to create more abstract concepts.

Cultural Evolution

The final issue we examine is that of cultural evolution. Lexical concepts are culturally embedded, and thus, we must briefly look at the role of cultural evolution in providing the conceptual backdrop for the emergence of semantic knowledge.

Consider the evolution of the concept of Money. This concept is one that has been evolving for over 3000 years. Weatherford (1998) identified a number of key mutations in the development of how we conceptualize Money. The first was the invention of coins in Anatolia over 3000 years ago. This development gave rise to the monetary economies that underpinned the classical Greek and Roman civilizations. The second was the development of family-owned credit banks in Renaissance Italy. This development gave rise to capitalist market economies that replaced earlier feudal societies throughout Europe, and

the period in which European countries expanded to become global economic powers. The process whereby cultural artifacts or cultural practice undergoes cumulative evolution, resulting in modification or improvement has been dubbed the 'ratchet effect' (Tomasello, 1999). Thus, an important aspect of the evolution of semantic knowledge involves the development and evolution of cultural knowledge.

Bibliography

- Barsalou L (1992). *Cognitive psychology*. Hillsdale: Lawrence Erlbaum.
- Barsalou L (1999). 'Perceptual symbol systems.' *Behavioral and Brain Sciences* 22, 577–660.
- Byrne R (1999). 'Human cognitive evolution.' In Corballis M & Lea S (eds.) *The descent of mind*. Oxford: Oxford University Press. 71–87.
- Deacon T (1997). *The symbolic species*. New York: Norton.
- Donald M (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.
- Donald M (1999). 'Human cognitive evolution.' In Corballis M & Lea S (eds.) *The descent of mind*. Oxford: Oxford University Press. 138–154.
- Evans V (2000). *The structure of time*. Amsterdam: John Benjamins.
- Fauconnier G & Turner M (2002). *The way we think*. New York: Basic Books.
- Gentner D (1981). 'Some interesting differences between verbs and nouns.' *Cognition and Brain Theory* 4(2), 161–178.
- Gentner D & Boroditsky L (2001). 'Individuation, relativity and early word learning.' In Bowerman M & Levinson S (eds.) *Language acquisition and conceptual development*. Cambridge: Cambridge University Press. 215–256.
- Grady J (1997). 'Theories are buildings Revisited.' *Cognitive Linguistics* 8(4), 267–259.
- Jackendoff R (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff R (1992). *Languages of the mind*. Cambridge, MA: MIT Press.
- Jackendoff R (2002). *Foundations of language*. Oxford: Oxford University Press.
- King J, Rumbaugh D & Savage-Rumbaugh S (1999). 'Perception as personality traits and semantic learning in evolving hominids.' In Corballis M & Lea S (eds.) *The descent of mind*. Oxford: Oxford University Press. 98–115.
- Lakoff G (1987). *Women, fire and dangerous things*. Chicago: Chicago University Press.
- Lakoff G & Johnson M (1980). *Metaphors we live by*. Chicago: Chicago University Press.
- Lakoff G & Johnson M (1999). *Philosophy in the flesh*. New York: Basic Books.
- Langacker R (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.

- Levelt W (1989). *Speaking*. Cambridge, MA: MIT Press.
- Mandler J (2004). *The foundations of mind*. Oxford: Oxford University Press.
- Savage-Rumbaugh S (1994). 'Hominid evolution: looking to modern apes for clues.' In Quiatt D & Itani J (eds.) *Hominid culture in primate perspective*. Boulder: University of Colorado Press. 7–49.
- Sperber D & Wilson D (1995). *Relevance* (2nd edn.). Oxford: Blackwell.
- Tomasello M (1999). *Cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Tomasello M (2003). *Constructing a language*. Cambridge, MA: Harvard University Press.
- Traugott E-C & Dasher R (2002). *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Turner M & Fauconnier G (1995). 'Conceptual integration and formal expression.' *Metaphor and Symbolic Activity* 10(3), 183–203.
- Tyler A & Evans V (2003). *The semantics of English prepositions*. Cambridge: Cambridge University Press.
- Varela F, Thompson E & Rosch E (1991). *The embodied mind*. Cambridge, MA: MIT Press.
- Weatherford J (1998). *The history of money*. New York: Three Rivers Press.
- Whiten A (1999). 'Human cognitive evolution.' In Corballis M & Lea S (eds.) *The descent of mind*. Oxford: Oxford University Press. 173–193.
- Ziemke T (2003). 'What's this thing called embodiment?' *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Hillsdale: Lawrence Erlbaum.

Evolution of Syntax

B Clark, Middlesex University, London, UK

R Jackendoff, Brandeis University, Waltham, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

After more than a century of more or less avoiding the subject, linguists have become much more interested in questions about evolution in recent years. Arguably, this interest was kick-started by an important paper by Steven Pinker and Paul Bloom (Pinker and Bloom, 1990). The number of articles in this encyclopedia that are concerned with matters of evolution can be seen as evidence of this increased interest (see cross-references below). Jackendoff (2002: 231–232) suggests two reasons for the increased interest. One is that we now understand much more about evolutionary principles in general and about human origins in particular than we did a century ago. The other is that increased interest in evolution in general inevitably leads to increased interest in the evolution of language in particular and, as Bickerton suggests (Calvin and Bickerton, 2000), it is important for linguists to be involved in this discussion. At the same time, there is some ambivalence about this topic. Linguists are wary of debates on topics where the data is so limited and the discussion therefore has to be quite speculative. Despite this, there have been a number of interesting discussions in recent years and significant progress has been made in this area. There has also been a significant increase in work that considers how syntax in particular might have evolved. This essay begins by considering some general questions about the nature and evolution of language before considering questions about the evolution of syntax, and some of the answers that have been proposed. The two sets

of questions are closely linked in ways that echo the connections between studies of language in general and studies of syntax in general.

Language

Before working out how a particular trait evolved, we first need to know what that trait is. There has been considerable disagreement about what 'language' is in the past and there are currently several different views (see the articles in this encyclopedia on **Linguistics: Approaches**). Language can be viewed as a social or a cultural phenomenon as well as a psychological one. But any social or cultural phenomenon must be represented mentally, and psychological systems and processes must be involved at some stage in the spread of any phenomenon through a culture.

Even when seen as a psychological phenomenon, there is disagreement about what kind of thing language is, and different assumptions about this will have implications for theories of language evolution. To take one example, Hauser *et al.* (2002) claim that recursion is the one key feature of language seen as a mental faculty, while Pinker and Jackendoff (2005) see recursion as just one of several important properties of language. At a more general level, there is the question of how components of the language system such as phonology (the system of sounds), syntax (structure), and semantics (representations of meaning) are related. Could phonology exist independently of syntax? Could syntax exist independently of phonology? Could the use of symbols exist without the existence of structural representations? Even more generally, there is the question of how language relates to other parts of the mind. Fodor's 'language

of thought' hypothesis (Fodor, 1975) suggests that our system of mental representations is both representational (has a kind of semantics) and computational (has a kind of syntax). If this is so, then we need to explore how the language of thought relates to our knowledge of natural languages, including the relationship between the syntax of the language of thought (which most theorists would assume to be universal) and the syntax of particular languages.

Evolution of Language

One of the most frequently discussed questions about the evolution of language concerns the speed at which it occurred. Was it a slow, incremental process where each stage developed independently after the previous one? Or did language emerge suddenly as something dramatically different from what had gone before? Another question concerns the nature of the evolutionary process. Pinker and Bloom (1990) suggest that language is a complex biological adaptation that evolved by natural selection, while other researchers have considered evolution as a cultural phenomenon rather than a biological one. Pinker and Bloom's work is partly motivated as a response to Chomsky's (1975) skepticism about an evolutionary account of language based on natural selection. While many researchers agree that a number of preadaptations made the development of language possible, there is disagreement on whether language emerged as an adaptation (so that language is now being used for the purpose for which it evolved) or, following Gould's (1991) terminology, an exaptation (where a trait that evolved for one purpose is then co-opted for another purpose). An exaptation can be either an adaptation that has been co-opted for a purpose different from that for which it originally evolved (such as feathers, which originally evolved because they helped keep birds warm but later turned out to be useful for flight) or side effects of adaptations, which then turn out to be useful in a different way. The latter are referred to as 'spandrels,' a term taken from architecture where it describes the spaces created as a side effect of putting two arches next to each other. Gould suggests that language could be a spandrel in the sense that it is a by-product of increased brain size that was beneficial for independent reasons. This view seems to be supported by Chomsky in one of his most well-known comments on evolution:

We know very little about what happens when 10^{10} neurons are crammed into something the size of a basketball, with further conditions imposed by the specific manner in which this system developed over time. It would be a serious error to suppose that all properties, or the interesting properties of the structures that

evolved, can be 'explained' in terms of natural selection. (Chomsky, 1975: 59)

Another general question concerns the relationship between the different features that together constitute human language. One important problem in developing an account of the evolution of language has to do with the perceived complexity of language and the large number of separate phenomena that together are involved in linguistic behavior: at least phonology, syntax, and semantics for the system itself, and at least intentionality and metarepresentational ability for the ability to communicate with it. While it is easy to see that full human language is likely to deliver an evolutionary advantage and so be favored by natural selection when it happens to emerge, it is not so obvious that any of these phenomena alone would deliver an evolutionary advantage. So it seems unlikely that a situation could arise where what we know as language could emerge and then be selected. This may be one reason for Chomsky's reluctance to entertain the possibility of an account of language evolution based on natural selection.

Partly despite and partly in response to this problem, discussion has focused on particular components of language and there has been ongoing discussion of the likelihood of various orders of emergence. Lieberman (1984, 1991, 1998), for example, has suggested that the ability to speak preceded the development of full language. He suggests that we first evolved all of the physiological resources to make speech possible and that this then made possible the development of full language. Corballis (2002), in contrast, suggests that fully syntactic messages were conveyed by hand gestures along the lines of modern signed languages long before speech developed. Bickerton (2003) rejects Corballis's view, following Burling (2000) and Sperber and Origgi (2000) in assuming that prehumans had to be attempting some form of intention-recognition before any of the components of language could exist. While he suggests that symbolism might have preceded structure by "as much as two million years" (Bickerton, 2003: 81), he also suggests that syntax is central to language and that this was the crucial development which made language possible. Okanoya (2002) suggests that syntax might have developed before words so that semantic relationships emerged after syntactic structure. This is based on evidence from the study of finch song, where it seems that a rudimentary syntax evolved independently of meaning. Okanoya further suggests that there are links between syntactic abilities and sexual selection.

An account of the evolution of syntax will be located within this debate and will have to address the

question of how syntax and its evolution relate to other components of language and their evolution.

One major issue for all work on evolution is the nature of the evidence. In work on the evolution of language, there is a wide range of approaches from a wide range of disciplines using a wide range of evidence types. Areas that provide, or could conceivably provide, evidence include animal communication, anthropology, archeology, artificial intelligence, cognitive neuroscience, comparative neuroanatomy, computer modeling, ethology, formal language theory, language acquisition, language disorders, mathematics and mathematical modeling, philosophy, psychological experiments, as well as more familiar linguistic and conceptual evidence. But of course since language itself leaves no fossils, all evidence is indirect.

Syntax

As suggested above, we cannot but assume that syntax is part of language seen as a psychological phenomenon (even if syntax-like combinatorial structure is also involved in thought and other aspects of the mind). So an account of the evolution of syntax must be part of an account of the evolution of language seen as a psychological phenomenon. As well as answering questions about the relationship between syntax and other components of language, any account of how syntax evolved will need to make assumptions about the nature of syntax itself and how it is related to other parts of the mind. Different answers to these questions lead to quite different accounts of the evolution of syntax. At one extreme, we might suggest that syntactic structure (in the sense of NPs and VPs) is necessary for combinatorial thought, and therefore that syntax is a necessary precursor to the development of full human language. At another extreme, we might suggest that syntactic structure emerges after the development of other phenomena such as the ability to speak and the ability to understand noises or gestures as representing particular meanings. To some extent, this is a question of the development and status of combinatoriality. Combinatoriality is a property of phonological and semantic knowledge, as well as of syntax. This raises the question of whether it arose in one of these areas first before developing in the other areas.

Chomsky has suggested that the properties of phonological structure and of semantic representations each depend on properties of syntax but not *vice versa*. In his view, syntax is fundamental to our linguistic knowledge. Jackendoff (2002) terms this 'syntactocentrism' and argues that it should be rejected.

He proposes that language can be understood in terms of a 'parallel architecture' where phonology, syntax, and semantics are separate, equally generative, components that communicate with each other through 'interface' components. This question concerns us here because of the implications for an account of the evolution of syntax. Can we treat the evolution of syntax as separate from the evolution of phonology and semantics or do they depend on each other, mutually or in an ordered way (where the evolution of one preceded the evolution of the other)?

Recursion is a fundamental property of syntax, and an important part of an evolutionary account will concern how knowledge of recursive structures such as (1) evolved:

- (1) Keith said that Billy thought that this was the book that annoyed the linguist who surprised the psychologist when he said that recursion was the most important defining property of human language.

What makes this recursive is the fact that it contains constituents that dominate other instances of the same syntactic category (for example, the sentence beginning *Keith said that...* dominates the sentence beginning *Billy thought that...*)

It is not only linguistic expressions that demonstrate recursion. Our thoughts too are recursive, in that we can embed thoughts within other thoughts, entertaining propositions such as that expressed by (2):

- (2) Tom thinks that Mary said that Dan was wondering whether Frieda knew about the party.

It seems that we are the only creatures who have this ability to such an extent. Other primates can entertain thoughts about thoughts but only to a limited extent (see Tomasello and Hare, 2003), e.g., a chimp might be able to think that if he defends an empty container, another creature will think it is full; or recursive thoughts might be involved in a chimp working out a plan of action. This is, though, about as complex as such thoughts can get for chimps. Humans can entertain recursive thoughts quite easily and can also produce and understand linguistic expressions that are complicated in the same way. This leads to speculation that the evolution of syntax may have made possible both metarepresentation and also human language.

Evolution of Syntax

There has been a considerable amount of significant work on the evolution of syntax in recent years

(a large amount of this work can be found in the collections edited by Christiansen and Kirby, 2003; Hurford *et al.*, 1998; Knight *et al.*, 2000; Wray, 2002). While no general consensus or dominant approach has yet emerged, we can identify a number of questions and themes that researchers have addressed. As mentioned above, many of these concerns overlap or mirror questions about the evolution of language as a whole. It is not possible here to do justice to the range of positions and themes that have been explored so far, but we run through a number to give a flavor of the kinds of issues currently being explored. At the end of this section, we give a more detailed sketch of Jackendoff's (2002) sketch of a possible series of stages in the evolution of (syntax and) language.

Is Syntax Adaptive?

As with other components of language, there are different views about whether syntax is adaptive or not, and therefore whether it can be explained in terms of natural selection or not. Pinker (2003; Pinker and Bloom, 1990) and Bickerton (1981, 1990, 1998, 2002, 2003), among others, suggest that language and syntax are adaptive and so amenable to an account in terms of natural selection. Lightfoot (1991, 2000), in contrast, suggests that some properties of syntax are not adaptive and so they have evolved despite being dysfunctional. Part of the evidence is that individuals violate some of the constraints imposed by the syntactic system. Lightfoot argues that certain conditions are “dysfunctional, blocking expressions which are needed” (Lightfoot, 2000: 244). This means that they are spandrels and so could not have been selected for. He concludes that this is true of at least some syntactic conditions but does not rule out the possibility that the whole language system is a spandrel. Overall, a number of distinct positions have been taken with regard to the question of whether syntax, or language, is adaptive. The notion that syntax is adaptive is often linked to the observation that combinatoriality is a property not only of syntax but also of semantics, phonology, and thought.

Exaptation

Assuming a biological approach, one suggestion that has a significant number of supporters is that syntax is an exaptation from nonsyntactic structure. One possible line on this is that the mental representations involved in thought developed combinatorial structure and this was then taken over by the linguistic system. Alternatively, combinatoriality might have

originated in phonology or in speech (Carstairs-McCarthy, 1999).

Biological or Nonbiological Evolution

Related to this is the question of whether explanations are proposed in terms of biology, of culture, or as developments within the language system itself. As we have seen, Bickerton and Pinker propose a biological account. Tomasello (2003) shares one part of Bickerton's view, namely that it is the separate evolution of symbols and grammar that distinguishes human communication from that of other primates. But he rejects Pinker's and Bickerton's view that it is specifically language that is adaptive. He suggests instead that language emerged as one part of the larger process of the evolution of human culture. In his view, the emergence of a broader kind of social cognition enabled human culture, and human symbolic communication should be seen as a special case of that. Worden (2000) proposes a model with a constant biological endowment which, together with specific learning mechanisms, enables the **cultural** evolution of languages. Hurford (2003) proposes a combination of biological preadaptations and learning-based linguistic adaptations. Jackendoff (2002) leaves open the question of how much is biological and how much cultural.

The ‘Big Picture’ and Details of Syntax

Hurford (2000: 222–224) raises the interesting issue of the relationship between programmatic, ‘big-picture’ statements and “the degree of detailed knowledge that has now been accumulated about the syntax of languages.” He points out, for example, the striking contrast between the simple syntaxes of the emergent languages in computer models and the complexity of real languages, which is familiar to linguists. He suggests that the value of programmatic proposals follows from the fragmented and rapidly changing nature of contemporary syntactic theorizing. The ‘turmoil’ of the field, he suggests, means that syntactic theorizing “needs to start looking beyond its traditional horizons for explanatory principles of kinds that it has not previously considered” (Hurford, 2000: 223).

A related contrast is between programmatic proposals and more specific suggestions. One of the boldest specific suggestions so far has been Newmeyer's (2000) suggestion, based on evidence from typology as well other work on language evolution, that the earliest forms of human language had rigid SOV word order.

Effects of Particular Developments Within Theories of Syntax

It is also interesting to consider how particular developments in syntactic theorizing relate to theories of the evolution of syntax. The Minimalist Program, for example, has been seen by some theorists (e.g., Berwick, 1998) as making evolutionary accounts more plausible, while others, including Chomsky himself (1991a, 1991b) have suggested that language is ‘dysfunctional’ with regard to processing and so makes an evolutionary account less likely. These conclusions rest, of course, on assumptions about how particular syntactic theories interact with other domains, such as theories of processing.

When Did Syntax Emerge?

Many researchers suggest that syntax emerged relatively late (around 40 000–150 000 years ago) and relatively quickly, but Burling (2002) suggests that it may have been a more gradual process starting much earlier. His evidence for this comes from looking at early and late stages of child language acquisition. He suggests that the acquisition of syntax spans a relatively long time in the development of an individual child and that this weakens arguments based on analogy with child language acquisition that the evolutionary development of syntax was abrupt.

How Many Stages?

Related to this is the question of the stages involved in evolution. Bickerton (1990, 1998, 2002, 2003; Calvin and Bickerton, 2000; see also Bickerton’s article in this encyclopedia) has proposed that language evolved in two stages. The second step is the development of language as we know it now. The first step is the development of protolanguage. Protolanguage is a simpler system of communication than full human language where certain actions (sounds or gestures) represent certain meanings but without the full complexity that arises with the development of syntax. So protolanguage can be thought of as ‘modern language without syntax.’

There are two particularly interesting things about this proposal. First, it suggests a way around the logical problem caused by thinking of human language as one entity and so suggests how we can develop an incremental account of language evolution. Protolanguage would bring advantages and therefore ‘be worth selecting for its own sake.’ Second, Bickerton suggests that protolanguage is still present in the human brain and can be seen in a number of situations where language is disrupted or

not developed. Cases of disruption include the development of pidgin languages when distinct language groups are thrown together and need to find a way to communicate without a full shared language (Bickerton, 1981), cases such as that of ‘Genie’ (Curtiss, 1977), where an individual is denied the necessary environmental triggers until it is too late for her to acquire full language, and possibly cases of agrammatic aphasia, language disorders that affect syntactic processing. Cases where language has not developed include early child language and communication systems developed by apes when humans attempt to teach them language (Linden, 1974; Savage-Rumbaugh *et al.*, 1998). The suggestion is that protolanguage emerged, was useful and favored by natural selection and that this made possible the subsequent development of human language with complex syntax.

A Series of Stages

Bickerton (1990) initially saw the development from protolanguage to modern language as one great leap, but has since (Calvin and Bickerton, 2000) developed a more gradualist position. Jackendoff (2002) proposes a much more incremental story. He suggests that it is possible to reconstruct from modern human language a sequence of distinct innovations from primate calls, each of which is an improvement in communicative expressiveness and precision. He suggests, furthermore, not only that the earlier stages are present in the brain but also that their ‘fossils’ are present in the grammar itself, and so are available as evidence. A summary of the steps proposed is shown in **Figure 1**.

Working downward through **Figure 1**, the first step and the two steps on the left of the figure (use of symbols, open class of symbols, phonological combinatorial system) take us beyond primate call systems but do not come close to full modern language. The development of a phonological combinatorial system is an important step here. As the number of symbols becomes larger, it becomes harder to make them all memorable and discriminable. If the symbols were holistic, it would not be possible to keep even a thousand of them distinct in perception and memory. This problem is solved in modern language by building words up combinatorially from a few dozen smaller meaningless speech sounds, the phonemes.

The first step toward modern syntax is to concatenate more than one symbol into a single utterance, with the connection between them determined purely by context. Concatenating more than two symbols multiplies the number of possible meanings, though much depends on exactly which symbols are used.

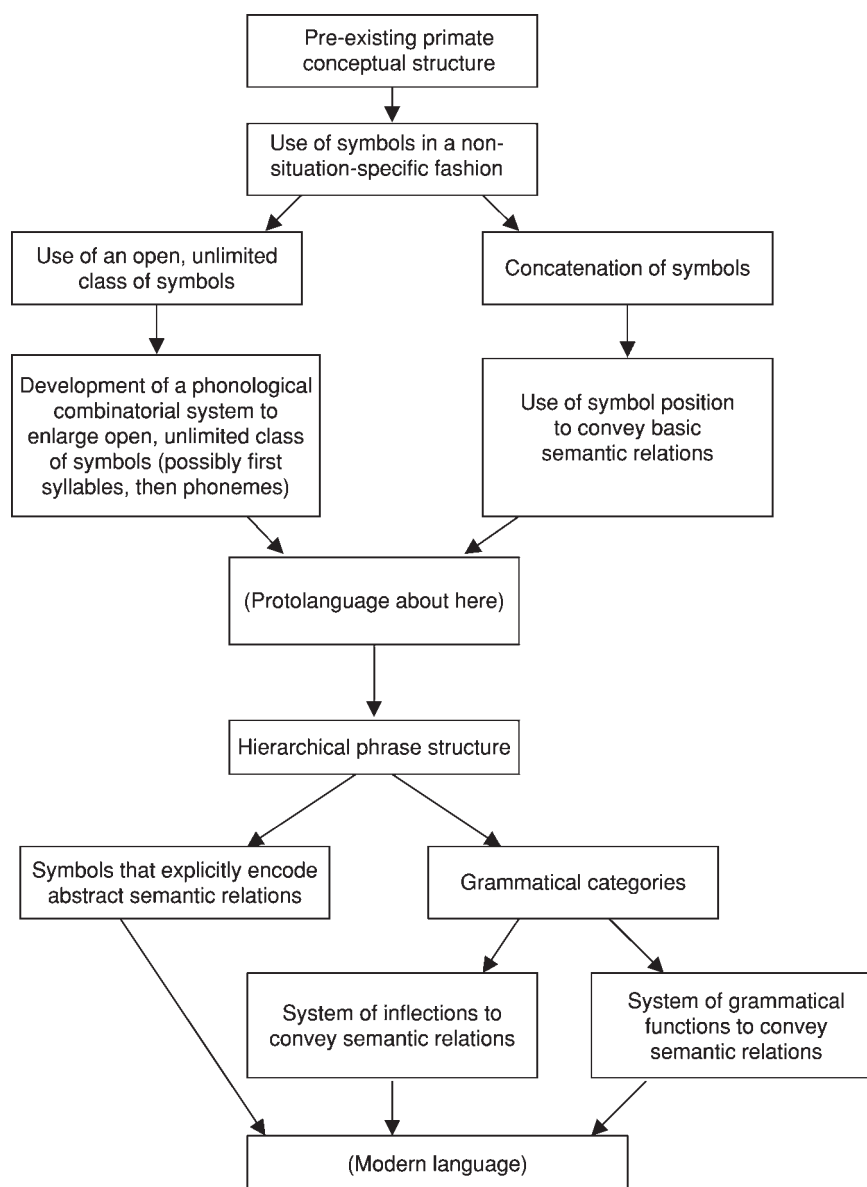


Figure 1 Summary of incremental evolutionary steps (from Jackendoff 2002: 238).

This is different from the combination of sounds just mentioned, in that meaningful symbols are combined to form larger utterances whose meanings are a function of the meanings of the symbols used. In ‘proto-phonology,’ in contrast, meaningless symbols are combined to form meaningful ones.

Concatenating symbols opens up many opportunities for enhancing expressive power and precision. Two important classes of innovations are orthogonal: using the linear order of concatenated symbols to express relations between them, and introducing new sorts of vocabulary item that convey relations explicitly. On this view, the former is a stage on the way to Bickerton’s protolanguage, while the latter

(discussed below) is a stage later than protolanguage on the way to modern language.

With just symbol concatenation, interpreting particular utterances depends very much on contextual inference. Pinker and Bloom (1990) suggest that this means that using principles of word order would be communicatively adaptive. Some principles of word order will narrow down the range of possibilities and lead to a system roughly equivalent to Bickerton’s protolanguage.

An important step in moving from protolanguage to modern language is to develop the means to represent relations not just between words but also between phrases. In a sentence such as *the careless boy*

lost his brother's apple, it is the entire phrase *the careless boy* that enters into a semantic relation with the verb *lost*. This collection of words functions as an elaborated version of the single word *boy*, the head of the phrase. More generally, a noun phrase is an elaborated version of a noun, an adjective phrase is an elaboration of an adjective and so on.

When headed phrases appear, then principles of word order can be elaborated into principles of phrase order. This means that the messages that can be conveyed are much more complex: not just *dog chase mouse*, for example, but also [*big dog with floppy ears and long scraggly tail*] *chase* [*little frightened mouse*]. It also makes possible hierarchical embedding such as [*the dog [that bit the cat [that chased the rat]]*]. This hierarchical embedding is one of the hallmarks of modern language but it is neither simple nor inevitable, even in a generative representation. It does not occur so relentlessly in phonological structure, for example.

This potential complexity raises new problems of communicability. With only three words in a sentence, the relations among them can be conveyed by simple word order plus pragmatics. But when sentences are longer and grouped into phrases, it becomes much more important to make phrase boundaries and the semantic relations among words explicit to the hearer. Language needs further devices than linear order and intonation in order to make semantic relations explicit.

One way to encode semantic relations among words and phrases is to invent words that express them. Relational words are pointless at the one-word stage but are useful once we have multiple-symbol utterances. The next step is the development of grammatical categories, and Jackendoff makes some suggestions about how the distinction between nouns and verbs might have arisen.

Two more changes are needed to move from this point to modern language: the development of morphology and the remaining aspects of syntax. Bickerton and many other linguists treat these as an integrated whole at the core of grammar, but Jackendoff suggests that we might see phrasal syntax and morphosyntax as independently evolved systems, each built on top of the system of protolanguage, each refining communication through its own expressive techniques.

On this speculative picture, then, modern language evolved from primate calls in a number of stages, each step having its own advantages over the step before. This suggests not only that the evolution of language can be seen as happening in a number of discrete steps but also that the syntactic component itself can be further broken down and that syntax itself may have

developed in a series of discrete steps. This leaves open the important question of which of these steps require biological evolution and which could be the consequence of cultural innovation.

Conclusion

Some linguists are reluctant to discuss a topic where the data is so remote. But the data of 'everyday' linguistics is also remote in that it is buried in the minds of language users at a level not accessible to consciousness. Recent work means that we now have a fuller understanding of evolution in general and linguists have been working on formulating and testing clearly stated hypotheses on the evolution of language, including the evolution of syntax. Continuing developments in the study of language and syntax have also helped linguists make significant progress in our attempt to understand the nature of syntax and language and how they evolved. While there is as yet no general consensus, there is a clear sense of an emerging interdisciplinary research program concerned with the evolution of language in general and with the evolution of syntax in particular.

See also: Evolution of Semantics; Linguistics: Approaches; Linguistics: Discipline of.

Bibliography

- Berwick R C (1998). 'Language evolution and the Minimalist Program: the origins of syntax.' In Hurford *et al.* (eds.). 320–340.
- Bickerton D (1981). *Roots of language*. Ann Arbor: Karoma.
- Bickerton D (1990). *Language and species*. Chicago: University of Chicago Press.
- Bickerton D (1998). 'Catastrophic evolution: the case for a single step from protolanguage to full human language.' In Hurford *et al.* (eds.). 341–358.
- Bickerton D (2002). 'Foraging versus social intelligence in the evolution of protolanguage.' In Wray (ed.). 207–225.
- Bickerton D (2003). Symbol and structure: a comprehensive framework for language evolution. In Christiansen & Kirby (eds.). 77–93.
- Burling R (2000). 'Comprehension, production and conventionalization in the origins of language.' In Knight *et al.* (eds.). 27–39.
- Burling R (2002). 'The slow growth of language in children.' In Wray (ed.). 297–310.
- Calvin W & Bickerton D (2000). *Lingua ex machina: reconciling Darwin and Chomsky with the human brain*. Cambridge: MIT Press.
- Carstairs-McCarthy A (1999). *The origins of complex language*. Oxford: Oxford University Press.

- Chomsky N (1975). *Reflections on language*. New York: Pantheon.
- Chomsky N (1991a). 'Linguistics and cognitive science: problems and mysteries.' In Kasher A (ed.) *The Chomskyan turn: generative linguistics, philosophy, mathematics and psychology*. Oxford: Basil Blackwell.
- Chomsky N (1991b). 'Some notes on economy of derivation and representation.' In Friedin R (ed.). *Principles and parameters in comparative grammar*. Cambridge: MIT Press. 417–454. Chomsky N (1995). Reprinted in *The Minimalist Program*. Cambridge: MIT Press. 129–166.
- Christiansen M H & Kirby S (eds.) (2003). *Language evolution*. Oxford: Oxford University Press.
- Corballis M C (2002). 'Did language evolve from manual gestures?' In Wray (ed.). 161–179.
- Curtiss S (1977). *Genie: a psycholinguistic study of a modern-day 'wild child.'* New York: Academic Press.
- Fodor J A (1975). *The language of thought*. Cambridge: Harvard University Press.
- Gould S J (1991). 'Exaptation: a crucial tool for evolutionary psychology.' *Journal of Social Issues* 47, 43–65.
- Hauser M D, Chomsky N & Fitch W T (2002). 'The faculty of language: what is it, who has it, and how did it evolve?' *Science* 298, 1569–1579.
- Hurford J R (2000). 'The emergence of syntax.' In Knight *et al.* (eds.). 219–230.
- Hurford J R (2003). 'The language mosaic and its evolution.' In Christiansen & Kirby (eds.). 38–57.
- Hurford J R, Studdert-Kennedy M & Knight C (eds.) (1998). *Approaches to the evolution of language: social and cognitive bases*. Cambridge: Cambridge University Press.
- Jackendoff R (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Knight C, Studdert-Kennedy M & Hurford J R (eds.) (2000). *The evolutionary emergence of language: social function and the origins of linguistic form*. Cambridge: Cambridge University Press.
- Lieberman P (1984). *The biology and evolution of language*. Cambridge: Harvard University Press.
- Lieberman P (1991). *Uniquely human: the evolution of speech, thought and selfless behavior*. Cambridge: Harvard University Press.
- Lieberman P (1998). *Eve spoke: human language and human evolution*. New York: Norton.
- Lightfoot D (1991). 'Subjacency and sex.' *Language and Communication* 11, 67–69.
- Lightfoot D (2000). 'The spandrels of the linguistic genotype.' In Knight *et al.* (ed.). 231–247.
- Linden E (1974). *Apes, men and language*. New York: Penguin.
- Newmeyer F J (2000). 'On the reconstruction of 'Proto-World' word order.' In Knight *et al.* (eds.). 372–390.
- Okanoya K (2002). 'Sexual display as a syntactical vehicle: the evolution of syntax in birdsong and human language through sexual selection.' In Wray (ed.). 46–63.
- Pinker S (2003). 'Language as an adaptation to the cognitive niche.' In Christiansen & Kirby (eds.). 16–37.
- Pinker S & Bloom P (1990). 'Natural language and natural selection.' *Behavioral and Brain Sciences* 13, 707–784.
- Pinker S & Jackendoff R (2005). 'The faculty of language: what's special about it?' *Cognition* 95(2), 201–236.
- Savage-Rumbaugh S, Shanker S & Taylor T (1998). *Apes, language and the human mind*. Oxford: Oxford University Press.
- Sperber D & Origgi G (2000). 'Evolution, communication and the proper function of the mind.' In Carruthers P & Chamberlain A (eds.) *Evolution and the human mind*. Cambridge: Cambridge University Press.
- Tomasello M (2003). 'On the different origins of symbols and grammar.' In Christiansen & Kirby (eds.). 94–110.
- Tomasello M, Call J & Hare B (2003). 'Chimpanzees understand psychological states: the question is which ones and to what extent.' *Trends in Cognitive Science* 7, 153–156.
- Worden R P (2000). 'Words, memes and language evolution.' In Knight *et al.* (eds.) 353–371.
- Wray A (ed.) (2002). *The transition to language*. Oxford: Oxford University Press.

Existence

B Caplan, University of Manitoba, Winnipeg, Canada

© 2006 Elsevier Ltd. All rights reserved.

What Existence Is

Existence is the property that is attributed to Uma Thurman in

- (1) Uma Thurman exists.

Perhaps existence is also attributed to some object in

- (2) There is an even prime.

There is a connection between existence and (objectual) quantification: what exists is exactly what our quantifiers quantify over, when our quantifiers are unrestricted. Sometimes our quantifiers are restricted so that they quantify over only some of the things that exist. For example, in

- (3) All the bottles of beer are in the fridge.

the quantifier "all the bottles of beer" is naturally interpreted so that it doesn't quantify over **all** of the bottles of beer in existence. But what exists is

not limited to what our quantifiers quantify over when they are restricted in one way or another. (In various free logics, variables need not be interpreted so as to have as values objects that exist. Sometimes a special predicate is introduced for 'exists' in these logics. Existence is not tied to quantification in these logics, although it might be tied to the special predicate.)

It seems that existence is a property that everything has: namely, the property *existing* or *being existent*. But various philosophers deny this for various reasons: some deny that existence is a property; others accept that existence is a property but deny that any objects have it (because only properties do); and still others accept that existence is a property but deny that all objects have it (because only some do).

The Hume-Kant View

The Scottish philosopher David Hume (1711–1776) and the German philosopher Immanuel Kant (1724–1804) denied that existence is a property. (It is often said that existence is not a predicate. This is at best a confused way of denying that existence is a property.) Let us call the view that existence is not a property the *Hume-Kant view*. One reason for holding the Hume-Kant view is that existence is supposedly not a property but rather a precondition for having properties. After all, how could something have any properties if it did not exist? But it is hard to see what a precondition is if it is not a property. For example, being human might be a precondition for being a movie star; and *being human* is a property.

Another reason for holding the Hume-Kant view is that to say that something has a property F and exists is supposedly not to say anything more than that something has F. For example,

- (4) Uma is a movie star and exists.

supposedly doesn't say anything more than

- (5) Uma is a movie star.

But if this is a good reason to deny that existence is a property, then it is also a good reason to deny that *being self-identical* or *being either round or not round* is a property. For if (4) doesn't say anything more than (5), then

- (6) Uma is a movie star and is self-identical.

and

- (7) Uma is a movie star and is either round or not round.

don't say anything more than (5) either. But it seems that *being self-identical* and *being either round or not*

round are perfectly respectable properties. For example, *being round* is a perfectly respectable property. And if negations and disjunctions of perfectly respectable properties are themselves perfectly respectable properties, then *being either round or not round* is also a perfectly respectable property.

The Frege-Russell View

Some philosophers who accept that existence is a property deny that everything has it, because they think that no *objects* have it; rather, they think that only *properties* have it. On this view, *existence* is not a (first-level) property of objects; rather, it is a (higher-level) property of properties. In particular, it is the property *being instantiated*. This is a view that was held by the German mathematician and philosopher Gottlob Frege (1848–1925) and, at least at one time, by the British philosopher Bertrand Russell (1872–1970). Let's call this view the *Frege-Russell view*.

One reason for holding the Frege-Russell view is that if existence were a property of objects, then it would not be possible to be mistaken in ascribing that property to an object. (By contrast, one can attribute the property *being instantiated* to the property *being a golden mountain*, say, even if that property is not instantiated.) But if this is a good reason to deny that existence is a property of objects, then it is also a good reason to deny that *being self-identical* or *being either round or not round* is a property of objects. For it is not possible to be mistaken in ascribing those properties to an object either. And yet they are perfectly respectable properties of objects.

Another reason for holding the Frege-Russell view comes from the problem of negative existentials. A *negative existential* is a sentence like

- (8) The golden mountain doesn't exist.

which seems to say of some object that it doesn't exist. For example, (8) seems to say, of the object that "the golden mountain" refers to, that it doesn't exist. Either "the golden mountain" refers to something or it doesn't. On the one hand, if "the golden mountain" doesn't refer to anything, then it seems that (8) doesn't say anything about anything. On the other hand, if "the golden mountain" does refer to something, then it seems that it must refer to something that exists, in which case (8) says, of something that does exist, that it doesn't exist. Either way, it seems that (8) can't be true. But (8) seems true; hence the problem.

The Frege-Russell view offers a straightforward solution to the problem of negative existentials. On the Frege-Russell view, (8) says, of the property *being the golden mountain*, that it does not have the

property *being instantiated*. And it is true that the property *being the golden mountain* does not have the property *being instantiated*. So, on the Frege-Russell view, (8) is true, as desired. (Russell's treatment of definite descriptions like "the golden mountain" is actually more complicated. (see **Descriptions, Definite and Indefinite: Philosophical Aspects**). One might worry that even if Russell's treatment solved the problem of negative existentials for sentences like (8), it wouldn't solve the problem of negative existentials for sentences like

(9) Santa Claus doesn't exist.

which contain names rather than definite descriptions (see **Proper Names: Philosophical Aspects**).

One problem with the Frege-Russell view is that (8) doesn't seem to say the same thing as

(10) The property *being the golden mountain* doesn't have the property *being instantiated*.

Similarly,

(11) If the golden mountain were to exist and if cows were to fly, then just as they would have the property *being able to fly*, it would have the property *being golden*.

seems true, and it doesn't seem to say the same thing as

(12) If the property *being the golden mountain* were to have the property *being instantiated* and if cows were to fly, then just as they would have the property *being able to fly*, it would be instantiated by something that has the property *being golden*.

Another problem with the Frege-Russell view is that the property *being instantiated* doesn't seem to be fundamental in the right sort of way. It seems that facts about which properties have the property *being instantiated* depend on quantificational facts. For example, it seems that the property *being a movie star* has the property *being instantiated* only because some object (Uma, say) instantiates the property *being a movie star*. But it seems that objects (Uma, say) can instantiate properties (*being a movie star*, say) only if they exist. So if it is to be instantiated, then the property *being instantiated* seems to require that some objects exist and hence that, contrary to the Frege-Russell view, existence be a property that at least some objects have.

The Meinong-Russell View

Some philosophers who accept that existence is a property deny that everything has it, because they think that some, but not all, objects have it. At one

time, Russell thought that there is a broad ontological property that everything has; but he thought that this property is *being* (or *subsisting*), not *existing*. On this view, the golden mountain, the round square, numbers, sets, tables, and chairs have being; but only tables and chairs (and other objects that are located in space and time) exist. The Austrian philosopher Alexius Meinong (1853–1920) held a similar view. He thought that there is a broad ontological property that everything has; but he thought that this property is *being an object*, not *being* or *existing*. On this view, the golden mountain, the round square, numbers, sets, tables, and chairs are objects; but of these, only numbers, sets, tables, and chairs have being. (And only tables, chairs, and other objects that are located in space and time exist.) Let's call this – the view that although there is a broad ontological property that everything has, only some objects exist – the *Meinong-Russell view*.

One reason for holding the Meinong-Russell view is that it offers a straightforward solution to the problem of negative existentials. On the Meinong-Russell view, (8) says, of the object "the golden mountain" refers to, that it doesn't exist, and "the golden mountain" refers to an object that doesn't exist. So, on the Meinong-Russell view, (8) is true, as desired.

But the Meinong-Russell view doesn't solve parallel problems. A *negative subsistential* is a sentence like

(13) The golden mountain has no being.

that seems to say of some object that it has no being. Those who distinguish being and existence sometimes say that "there is" has to do with being, not existence. On this view,

(14) There is no golden mountain.

is also a negative subsistential. And a *negative objectual* is a sentence like

(15) The golden mountain isn't an object.

or

(16) No object is the golden mountain.

that seems to say of some object that it isn't an object. Speakers who have the intuition that (8) is true might also have the intuition that (13)–(16) are true. And if a solution to the problem of negative existentials should respect speakers' intuition that (8) is true, then one might think that a solution to the problem of negative subsistentials or negative objectuals should similarly respect speakers' intuition about (13)–(16). But on the Meinong-Russell view, (13) and (14) or at least (15) and (16) are false, because "the golden mountain" refers to an object that has being or at least is an object. (This argument might work best against those who say that (8) is false but (14) is true.) Solving the problem of negative existentials only at the cost of not solving the problem of

negative subsistentials or the problem of negative objectuals doesn't seem like much of a benefit. In addition, many dislike the Meinong-Russell view because, by saying that existence is what Russell (1903) once called "the prerogative of some only amongst beings," the view offends what Russell (1919) later described as "a robust sense of reality."

If one rejects the Hume-Kant view, the Frege-Russell view, and the Meinong-Russell view, one is left with the view that existence is a property that everything has. Although there is much to commend this view, those who hold it still have to solve the problem of negative existentials. This suggests that a solution to that problem will not come from views about existence. And once one had a solution to the problem of negative existentials (whatever that solution is and wherever it comes from), it seems that there would be little to prevent one from holding the view that existence is a property that everything has.

See also: Descriptions, Definite and Indefinite: Philosophical Aspects; Empty Names; Fictional Discourse: Philosophical Aspects; Negation: Philosophical Aspects; Nominalism; Objects, Properties, and Functions; Proper Names: Philosophical Aspects.

Bibliography

- Frege G (1884). *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Breslau: Koebner. Austin J L (trans.) (1950). *The foundations of arithmetic: a logico-mathematical enquiry into the concept of number* (2nd edn., 1980). Evanston, IL: Northwestern University Press.
- Frege G (1892). 'Über Begriff und Gegenstand.' *Vierteljahrsschrift für wissenschaftliche Philosophie* 16, 192–205. Black M (trans.) (1952). 'On concept and object.' In Geach P T Black M (eds.) (1952). *Translations from the philosophical writings of Gottlob Frege* (3rd edn., 1980). Oxford: Blackwell. 42–55. Reprinted in Frege G (1997). Beaney M (ed.). *The Frege reader*. Oxford: Blackwell. 181–193.
- Hume D (1740). *A treatise of human nature: being an attempt to introduce the experimental method of reasoning into moral subjects*. London: Noon. Reprinted in Norton D F & Norton M J (eds.) (2000). *A treatise of human nature. Oxford philosophical texts*. Oxford: Oxford University Press.
- Kant I (1781). *Kritik der reinen Vernunft*. (2nd edn., 1787). Riga: Hartknoch. Guyer P & Wood A W (trans.) (1998). *Critique of pure reason*. Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press.
- Meinong A (1904). 'Über Gegenstandstheorie.' In Meinong A (ed.) *Untersuchungen zur Gegenstandstheorie und Psychologie*. Leipzig: Barth. Levi I, Terrell D B & Chisholm R M (trans.) (1960). 'The theory of objects.' In Chisholm R M (ed.). *Realism and the background of phenomenology*. Glencoe, IL: Free Press. 76–117.
- Quine W V O (1948). 'On what there is.' *Review of Metaphysics* 2(5) (Sept.), 21–38. Reprinted in Quine W V O (1961). *From a logical point of view: nine logico-philosophical essays* (2nd edn., 1953). Cambridge, MA: Harvard University Press. 1–19.
- Russell B (1903). *The principles of mathematics*. Cambridge: Cambridge University Press.
- Russell B (1905). 'On denoting.' *Mind* 14(56), 479–493. Reprinted in Urquhart A (ed.) (1994). *The collected papers of Bertrand Russell, vol. 4: Foundations of logic, 1903–05*. New York, NY: Routledge. 414–427.
- Russell B (1918–1919). 'The philosophy of logical atomism.' *Monist* 28(4), (Oct. 1918): 495–527, 29(1) (Jan. 1919): 32–63; 29(2) (April 1919): 190–222; 29(3) (July 1919): 345–380. Reprinted in Slater J (ed.) (1986). *The collected papers of Bertrand Russell, vol. 8: The philosophy of logical atomism and other essays, 1914–19*. London: Allen & Unwin. 157–244.
- Russell B (1919). *Introduction to mathematical philosophy*. Muirhead Library of Philosophy. London: Allen & Unwin.
- Salmon N (1987). 'Existence.' In Tomberlin J E (ed.) *Philosophical perspectives, vol. 1: Metaphysics*. Atascadero, CA: Ridgeview. 49–108.

Expression Meaning versus Utterance/Speaker Meaning

A Bezuidenhout, University of South Carolina, Columbia, SC, USA

© 2006 Elsevier Ltd. All rights reserved.

When Mrs. Malaprop in Richard Sheridan's play *The Rivals* says to her niece Lydia Languish "don't attempt to extirpate yourself from the matter," she

means to say that her niece should not attempt to *extricate* herself from the matter. But that is not what 'extirpate' means in English (at least, it is not a meaning one would find listed under 'extirpate' in a good dictionary of English usage). Malapropisms of this sort are one way in which expression meaning (i.e., word or sentence meaning) can come apart from speaker meaning. Mrs. Malaprop has a

mistaken belief about what the words she is using mean in the language she is using. Slips of the tongue (e.g., saying 'pig vat' instead of 'big fat') represent another way in which expression and speaker meaning can come apart.

Gricean conversational implicatures represent another, much larger, class of cases in which these two kinds of meaning come apart. These are cases in which the speaker engages in some form of indirection, where, typically, the main conversational point is something implicitly communicated rather than explicitly expressed. In such cases, the speaker's words mean one thing, but the speaker is trying to convey another meaning, either in addition to the literal expression meaning or in place of it. An example of the former sort is when Mary replies to Peter's offer to take her to the movies that evening that she will be studying for an exam then. Mary has explicitly said that she will be studying, but has implicitly communicated that she is refusing Peter's invitation. Here both the explicit statement and the implicit refusal are intentionally communicated. The statement is intended to give Mary's reason for her refusal. An example of the latter sort is when Mary responds to Peter's refusal to help her when she is in need by saying 'You're a fine friend!' Here she is implicitly communicating that Peter is *not* a good friend. Her words 'fine friend' are being used sarcastically, and she does not intend to communicate what her words literally mean.

It should be mentioned that there are philosophers who think that even what is explicitly said (as opposed to implicitly communicated) can come apart from literal sentence meaning. These are cases where literal expression meaning must be pragmatically narrowed or broadened in order to arrive at what is explicitly communicated. Thus, when Mary says to the waiter at the restaurant that he should take her steak back because it is raw, she doesn't mean to say the steak is literally uncooked, but that it is too undercooked for her taste – a case of pragmatic broadening. Or when Mary tells her son that he is not going to die when he comes crying to her with a cut on his finger, she means to say that he is not going to die from that cut, not that he is never going to die – a case of pragmatic narrowing (see **Pragmatic Determinants of What Is Said**).

For some, utterance meaning is just a variety of speaker meaning. It is the meaning an expression has as used by a speaker in some conversational context. The hearer arrives at an understanding of utterance meaning by combining literal expression meaning with other contextually available information, including information about the speaker's communicative intentions. However, at least some

philosophers of language and linguists wish to draw a contrast between utterance and speaker meaning. Levinson (1987, 1995, 2000) has argued for three levels of meaning. There is expression meaning, utterance meaning, and speaker meaning. Utterance meanings belong to a system of default meanings associated with certain expression types. These default meanings are distinct from literally encoded expression meanings. However, when a speaker utters an expression of this type in a normal context, she will have conveyed the default meaning, unless she either explicitly or implicitly cancels this meaning. Levinson identifies these default meanings with the class of conversational implicatures that Grice called generalized conversational implicatures. For instance, when Peter accuses Mary of having eaten all the cookies and Mary replies that she has eaten some of the cookies, she explicitly says that she has eaten some and possibly all of the cookies, she implicates in a generalized way that she has not eaten all of the cookies, and she implicates in a particularized way that she is not the guilty party. These three meanings correspond to Levinson's three levels of sentence, utterance, and speaker meaning, respectively.

The distinction between expression and speaker meaning has been invoked in many philosophical debates as a way of avoiding the postulation of multiple meanings for a single expression type. One well-known instance is Kripke's (1977) appeal to a distinction between speaker's reference and semantic reference of definite descriptions. Kripke appealed to this distinction in order to deny the semantic significance of what Donnellan (1966) called the referential use of such descriptions. Suppose Mary uses the description 'the man in the corner drinking a Martini,' intending to refer to Peter, but in fact Peter is drinking water, not Martini. Kripke argues that the so-called referential use of the description can be accounted for by appeal to what Mary meant to convey by the use of that expression, whereas what she actually said is determined by giving a Russellian analysis of the description. Since there is no unique Martini drinker in the corner (since, let us suppose, there is no Martini drinker there), what Mary has said is false, although what she meant to convey (her speaker meaning) may very well have been true.

There are differing views as to the relative priority of expression and speaker meaning. Some philosophers, such as Strawson (1950), have argued that it is not words and sentences by themselves that refer or express propositions. Rather, it is speakers who refer or express propositions by their uses of words and sentences, respectively. Salmon (2004) calls this the speech-act-centered conception of semantics and contrasts it with the view he favors, namely the

expression-centered conception. According to the latter conception, words and sentences have their semantic properties intrinsically, in the sense that one can talk about the referential and truth-conditional content of expressions without any knowledge of or appeal to the communicative intentions of users of those expressions.

Although defenders of the speech-act-centered conception are committed to denying that expressions have referential or truth-conditional content independently of speakers' communicative intentions, their view is compatible with the claim that expression types have aspects of meaning that are context invariant. These would correspond to Fregean 'senses' or 'modes-of-presentation' or (for demonstratives and indexicals) to Kaplanian 'characters.' Such nonreferential or nontruth-conditional aspects of meaning may be intrinsic in Salmon's sense. In other words, such meaning would be a property of expression types, independently of the intentions of the users of those expression types.

Some philosophers of language have denied the idea of intrinsic expression meaning independent of speaker meaning. For instance, Grice (1957) argued that expression meaning is reducible to speaker meaning. Grice was interested in nonnatural meaning (Meaning_{NN}), as opposed to the sort of natural meaning that a sign may have in virtue of naturally signaling or indicating some state of affairs. He argued that an utterance's nonnaturally meaning that *p* is simply a matter of a speaker's uttering an expression with a certain communicative intention. This would be a sort of 'one-off' meaning for that expression. However, that speaker may be disposed to utter an expression of this type whenever he wishes to convey a certain meaning. Thus, he might develop a habit of using that expression type that way. If this usage were then to spread to other members of his community, it would become a standardized usage, and that expression type would come to have a stable meaning independent of the intentions of any one speaker. But such a meaning would not be independent of the linguistic activities of the users of the expression type in general.

Another way that defenders of a speech-act-centered conception have challenged the idea of intrinsic

expression meaning is to argue with Searle (1983) that all meaning is relative to a nonintentional Background. A sentence only has truth-conditions relative to some assumed Background. This Background can never be made fully explicit, because at bottom it consists in a set of abilities, practices, and ways of acting that are nonintentional. Although Searle, unlike Grice, is not suggesting that expression meaning depends ultimately on the communicative intentions of speakers, he is arguing that expression meaning depends on a certain sort of human activity, and so this conception is antithetical to the idea of intrinsic expression meaning.

See also: Character versus Content; Context Principle; Conventions in Language; Descriptions, Definite and Indefinite: Philosophical Aspects; Intention and Semantics; Natural versus Nonnatural Meaning; Pragmatic Determinants of What Is Said; Referential versus Attributive; Semantics–Pragmatics Boundary; Sense and Reference: Philosophical Aspects; Speech Acts; Truth Conditional Semantics and Meaning.

Bibliography

- Donnellan K (1966). 'Reference and definite descriptions.' *Philosophical Review* 75, 281–304.
- Grice P (1957). 'Meaning.' *Philosophical Review* 66, 377–388.
- Kripke S (1977). 'Speaker's reference and semantic reference.' *Midwest Studies in Philosophy* 2, 255–276.
- Levinson S (1987). 'Minimization and conversational inference.' In Verschueren J & Bertuccelli-Papi M (eds.) *The pragmatic perspective*. Amsterdam: John Benjamins. 61–129.
- Levinson S (1995). 'Three levels of meaning.' In Palmer FR (ed.) *Grammar and meaning*. Cambridge: Cambridge University Press. 90–115.
- Levinson S (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Salmon N (2004). 'The good, the bad and the ugly.' In Reimer M & Bezuidenhout A (eds.) *Descriptions and beyond*. Oxford: Oxford University Press. 230–260.
- Searle J (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Strawson P (1950). 'On referring.' *Mind* 59, 320–344.

Expressive Power of Language

D Gil, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

© 2006 Elsevier Ltd. All rights reserved.

The expressive power of a language is its ability to convey meanings of various kinds. Most often, the term is used in comparative or contrastive contexts, for example, to claim that one language is of greater expressive power than another in a certain semantic domain.

At one time, it was believed that the classical languages of Western civilization enjoyed much greater expressive power than the so-called primitive languages of Africa, Asia, Oceania, and the Americas. However, when linguists began familiarizing themselves with more distant and exotic languages, it became clear that they were anything but impoverished in their expressive power. Today, it is generally accepted that all languages are endowed with expressive power of similar order of magnitude.

Nevertheless, within particular domains, languages may still vary substantially with respect to their expressive power. Perhaps the most celebrated example of crosslinguistic variation in expressive power, the Eskimos' purported plethora of 'snow words,' has recently been called into question; however, numerous other examples are empirically well-supported. In order to compare the expressive power of languages, it is necessary to consider the forms that may bear meanings, the meanings that may be borne by those forms, and the nature of the form-meaning relationship.

The forms that may bear meanings are linguistic units of various size: morphemes, words, phrases, clauses, sentences, and texts. Alongside these, intonation contours may also be associated with meanings. In addition, smaller, phonological units may bear meanings in what is known as sound-symbolism.

The meanings that are borne by linguistic forms may be broadly characterized as either conceptual or affective. Conceptual meanings are based on the traditional notions of semantics, such as propositional content, modality, and reference, and may be further classified into various semantic domains, including thematic roles, tense/aspect, quantification, kinship terms, color terms, weather terms, and so forth. In contrast, affective meanings involve states such as tension, happiness, arousal, and their like.

The form-meaning relationship, that is to say the association of linguistic forms with their particular meanings, is the basis of the expressive power of language. For any given meaning *M*, the following

three cases may, potentially, be distinguished: (a) *M* is expressible in all languages; (b) *M* is expressible in no languages; and (c) *M* is expressible in some but not all languages. However, the above taxonomy faces a number of serious problems.

One difficulty is empirical: without access to each and every one of the world's thousands of languages, it is an extremely risky proposition to make any universal generalizations to the effect that a certain meaning may be expressible in all languages, or in none. In lieu of a much larger data base than is currently available, all such claims must therefore be treated as conjectures of various degrees of plausibility.

Another issue pertains specifically to the meanings purportedly expressible in no languages. Some scholars view language as system of societal conventions, and meanings, in particular, as essentially public entities: for such scholars, the concept of an unexpressible meaning is therefore an oxymoron. More generally, by their very definition, meanings expressible in no language can never be referred to, even indirectly, as in expressions such as *the meaning of Mona Lisa's smile*: the very notion of unexpressible meaning is thus paradoxical.

Nevertheless, prophets, poets, and others have reported undergoing mystical experiences beyond the realm of linguistic expression; similar claims are also made by persons who have entered altered states of consciousness, for example, with hallucinogenic drugs. Indeed, the logical positivists maintained that natural languages are incapable of expressing scientific concepts with sufficient precision; this postulate was their motivation for proposing an 'artificial language of science.' More far reaching, cognitive scientists have suggested that the human mind is innately incapable of grasping certain kinds of knowledge; such knowledge would, *a fortiori*, be unexpressible in any language. In view of such claims, it would therefore seem advisable to at least entertain the possibility that there may exist meanings expressible in no languages, in spite of the problems inherent therein.

A further, much-debated issue is that of crosslinguistic synonymy. Some linguists question whether linguistic forms in different languages are ever capable of bearing the exact same meaning. Each and every linguistic form is part of a language system; hence, it is argued, linguistic forms belonging to different language systems can never be equivalent. The issue of crosslinguistic synonymy is of course crucial to the process of translation: if complete crosslinguistic synonymy is not attainable, neither is the 'perfect translation.' Nevertheless, for many purposes, it would seem advantageous to abstract away from

such subtleties, in order to posit crosslinguistic synonymy, when equivalence of meaning, up to a certain point, can be seen to obtain.

A particular problem arises when supposed crosslinguistic synonyms are constructed from linguistic forms of different types, or 'sizes.' Consider, for example, the English expression *boy* and its nearest equivalents in Hebrew, *yeled*, and in Tagalog, *batang lalaki*. Whereas English has a monomorphemic word, Hebrew possesses a bimorphemic word consisting of root *y-l-d* 'child' plus masculine singular inflection *-e-e-*, while Tagalog makes use of two words, *batang* 'child' plus a grammatical linker, and *lalaki* 'male.' It is at least plausible to suggest that the meanings associated with *boy*, *yeled*, and *batang lalaki* reflect their different structures. Specifically, whereas a semantic representation of *boy* might involve a primitive concept Boy, that of *yeled* might be obtained by application of a masculine feature to Child, while that of *batang lalaki* might be constructed compositionally, by modifying Child with Male.

It would therefore seem crucial to distinguish between the different types of linguistic forms that may bear meanings: segments, morphemes, words, phrases, clauses, sentences, texts, intonation contours, and so forth. Specifically, for any meaning M, and any type of linguistic form T, the following three cases may be distinguished: (a) M is expressible by a form of type T in all languages; (b) M is expressible by a form of type T in no languages; and (c) M is expressible by a form of type T in some but not all languages. Following are some examples of each of these three cases.

Taking T to be morpheme, and examining the semantic domain of cardinal numerals, One is possibly expressible by a single morpheme in all languages, whereas One Hundred And Twenty Three is almost certainly expressible by a single morpheme in no languages. In between these two extremes, Eleven is expressible by a single morpheme in English *eleven*, but not in Malay, where *seblas* consists of *se-* 'one' plus *-blas* 'over ten'; while One Hundred Thousand is expressible by a single morpheme in Hindi *laakh*, but not in English.

Letting T stand for word and generalizing the semantic domain to that of quantification, One is possibly expressible by a single word in all languages, whereas At Most Four Or More Than Seven is almost certainly expressible by a single word in no languages. In between, the quantifier No is expressible by a single word in English *no* (as in *no languages*), but not in Hebrew, where it would be paraphrased with a negative polarity marker *af* preceding the noun plus a negation *lo* in construction with the verb; while

Only One is expressible by a single word in Tagalog *iisa*, but not in English.

Taking T to denote clause, simple propositions such as John Came are perhaps expressible by a single clause in all languages, whereas a conditional linking two events such as If John Comes, Bill Will Go most probably cannot be expressed monoclausally in any language. However, propositions involving conjoined participants such as John And Bill Came can be expressed by a single clause in English, *John and Bill came*, but not in the Amerindian Yuman language Maricopa, where its nearest equivalent, *John-f Bill uðaavk vʔaawk*, contains an embedded clause *John-f Bill uðaavk* 'John accompanied Bill.' Conversely, propositions involving indirect causation such as John Caused Bill To Write A Letter can be expressed monoclausally in Hindi, as in *John-ne Bill-se khat likh-waayaa*, making use of a single verb *likhwaayaa* 'write' inflected for indirect causation, but not in English, where, in the nearest equivalent, *John caused Bill to write a letter*, two verbs, *caused* and *write*, head two different clauses.

Letting T represent intonation contour, emphasis is probably expressible via intonation in all languages, while a cardinal numeral such as Seven is almost certainly expressible through intonation in no languages. However, yes/no questions can be formed with intonation in English, but not, allegedly, in Thai, whereas 'a variety of exclamation colorings,' including 'the lack of concern on the part of the speaker for the outcome of his utterance' can be expressed intonationally in Vietnamese, but not in English.

Finally, taking T to denote phonological feature, harshness and aggression are possibly expressible, through sound symbolism, by the feature [obstruent] in all languages, while a cardinal numeral such as Seven is almost certainly expressible with a phonological feature in no languages. However, since the association of phonological features with meanings is based on universal cognitive principles, falling outside the domain of conventionalized and possibly language-specific grammatical rules, there are probably no cases of meanings expressible by phonological features in some but not all languages.

In those cases where a meaning M is expressible by a form of type T in some but not all languages, the availability of M as a meaning of a Type T form may be correlated with various other linguistic properties. Such correlations can be represented as implicational universals of the following form: (a) For any language L, if M is available as a meaning of a Type T form in L, then L has property P; and (b) For any language L, if L has property P, then M is available as a meaning of a Type T form in L. At present, however, not enough is

known about patterns of crosslinguistic variation in expressive power to conclusively support a significant body of such universals. Nevertheless, some work in linguistic typology has pointed towards possible correlations between expressive power and various other linguistic properties.

For example, in phonological typology, it has been suggested that intonation is associated with less expressive power in tonal languages than in their nontonal counterparts. Similarly, in morphological typology, the word is generally endowed with less expressive power in isolating languages than in synthetic ones, and less expressive power in synthetic languages than in polysynthetic ones, such as the Australian language Mayali, in which a single word *gayauganjngunihmiwage* can mean 'the child crawls along eating meat.' Analogously, in syntactic typology, the sentence is generally equipped with less expressive power in English and other European languages than in several Papuan languages of the so-called clause chaining type, in which a single sentence can narrate a series of events that, in most other languages, would require a lengthy sequence of sentences.

While some facts and generalizations about the expressive power of language can be accounted for terms of linguistic typology, others would seem to result from extralinguistic factors, while yet others would appear to be immune to either kind of explanation,

instead reflecting the arbitrariness of language and linguistic structure.

See also: Grammatical Meaning.

Bibliography

- Berlin B & Kay P (1969). *Basic color terms: their universality and evolution*. Berkeley: University of California Press.
- Bolinger D (1989). *Intonation and its uses: melody in grammar and discourse*. London: Edward Arnold.
- Fónagy I (1983). *La Vive Voix, Essais de Psycho-Phonétique*. Paris: Payot.
- Keenan E L (1975). 'Logical expressive power and syntactic variation in natural languages.' In Keenan E L (ed.) *Formal semantics of natural language*. Cambridge: Cambridge University Press. 406–422.
- Keenan E L & Stavi J (1986). 'A semantic characterization of natural language determiners.' *Linguistics and Philosophy* 16, 253–326.
- Keffer M & van der Auwera J (eds.) (1992). *Meaning and grammar: cross-linguistic perspectives*. Berlin: Mouton de Gruyter.
- Talmy L (1985). 'Lexicalization patterns: semantic structure in lexical forms.' In Shopen T (ed.) *Language Typology and Syntactic Description 3: Grammatical categories and the lexicon*. Cambridge: Cambridge University Press. 57–149.
- Zaefferer D (ed.) (1991). *Semantic universals and universal semantics*. Berlin: Foris Press.

Extensionality and Intensionality

N Oldager, Technical University of Denmark, Lyngby, Denmark

© 2006 Elsevier Ltd. All rights reserved.

A sentence is extensional if its expressions can be substituted with expressions that have the same denotation (reference) without altering the truth value of the sentence. A sentence that is not extensional is intensional. A language is extensional if every sentence of it is extensional. Otherwise, the language is intensional.

The following sentence is then intensional:

George IV wished to know whether Scott was the author of *Waverly*.

As Scott was in fact the author of *Waverly*, 'Scott' and 'the author of *Waverly*' are co-denotational. However, if we substitute one for the other, we get

George IV wished to know whether Scott was Scott,

which, unlike the former, can hardly be taken as true. Because natural language contains intensional sentences, natural language is intensional.

A context in which co-denotational expressions cannot be substituted is known as an indirect context (or oblique, opaque or intensional context), and a context of extensional expressions is a direct context. Sentences involving propositional attitudes, intentions, quotations, temporal designation, and modalities give rise to indirect contexts. Another example of intensionality:

Nine necessarily exceeds seven.

Nine is the number of the planets.

The number of the planets necessarily exceeds seven.

Although the first two sentences are true, the third is not because it is only a contingent astronomical fact

and not a necessary truth that the number of planets exceeds seven – it is possible that there were only seven planets.

Semantical Aspects of Extensionality and Intensionality

Issues concerned with extensionality and intensionality have been cardinal motivations behind the development of important semantical theories. Reducing extensionality and intensionality to technical conditions regarding substitutivity of expressions is accordingly a crude simplification.

Though extensionality and intensionality can be traced as far back as ancient Greek philosophy, the first major contribution to the subject was Gottlob Frege's *Über Sinn und Bedeutung* (Frege, 1892). Note, there are different translations of the title words of this work: *Sinn* is translated as 'sense' but *Bedeutung* is translated as either 'denotation', 'reference', or 'nominatum' ('meaning' has actually been used for both *Sinn* and *Bedeutung*). Following Bertrand Russell and Alonzo Church, *Bedeutung* will be identified with 'denotation'.

Although Frege does not spell it out in detail, he maintains that semantics is compositional, such that the semantics of a sentence is determined by the semantics of its parts. To illustrate his theory, assume, to begin with, that semantics is purely referential, that is, assume that the semantics of an expression is what the expression denotes. This seems plausible; for example,

Paris is beautiful

asserts that what 'Paris' denotes, i.e., the actual capital of France, has the property of being beautiful. It is not the string of symbols 'Paris' or one's idea of Paris, whatever that may be, which is beautiful.

However, things are more complicated than this. There are aspects of natural language semantics that cannot be explained by resorting to the notion of denotation, Frege argues. He illustrates this by the following puzzle. Suppose a and b are names for some objects and that

$$a = b$$

is true, hence the expressions ' a ' and ' b ' have the same denotation.

Frege then recognizes a difference between this identity and an identity such as $a = a$. The latter is trivially true (analytically true), whereas the former may contain useful information. For instance, in a criminal investigation a discovery such as "the burglar is the suspect" could be decisive, whereas "the burglar is the burglar" is useless. The important question is

then: What is the source of the difference between $a = a$ and $a = b$? As semantics is compositional, the difference must be due to a difference between the semantics of the expressions ' a ' and ' b '. But by assumption, ' a ' and ' b ' have the same semantics because they have the same denotation. In other words, referential semantics must be rejected because it cannot explain the difference between the identities.

Frege's famous solution is to acknowledge that ' a ' and ' b ' refer to their denotation in different ways. Consider, for example, the expressions 'morning star' and 'evening star.' Both denote the planet Venus, so *morning star* = *evening star* is true, but they refer to Venus in different ways: One refers to a heavenly object seen in the morning, the other to a heavenly object seen in the evening. Frege says 'morning star' and 'evening star' have different senses. The puzzle about identity can now be solved by noting that $a = a$ and $a = b$ express different senses. Frege was inspired by mathematics when he developed this theory. Consider the two expressions ' $1 + 3$ ' and ' $2 * 2$ '. Both are equal to, i.e., denote, the number 4, however, their way of referring to four differs because the calculations for obtaining the result differ. Hence, the expressions have different senses.

To solve the puzzle, Frege accordingly introduces two semantical concepts, denotation and sense. Each expression – including proper names and sentences – is then assumed to have both a denotation as well as a sense, although he recognizes that exceptions may occur. Frege never precisely described what senses are, but he explained that the sense of an expression contains its mode of presentation (its way of referring), that expressions express their senses, and senses are something we grasp. Moreover, senses are distinguished from ideas (subjective thinking), meaning they are objective. The notion of sense may appear unfamiliar and it may not be clear why it is seminal. However, expressions such as 'morning star' or 'Paris' are signs, and signs are characterized by their ability to refer to something. Sense addresses this fundamental feature – the referential capacity of expressions. So, it is natural to discern between expressions that refer to the same thing.

There is a close relationship between Frege's theory and the earlier definition of extensionality and intensionality. Frege can now provide an explanation for failure of substitutivity in indirect contexts. In indirect contexts we are talking about the senses of the expressions occurring therein. When we say

John believes that the morning star is the evening star

we are not talking about Venus, Frege argues, but about different senses that determine Venus. In

indirect contexts the semantics become the sense. Thus, when we substitute two co-denotational expressions that have different senses in indirect contexts, we obtain different propositions that may have different truth values.

Frege's theory has been challenged by, among others, Bertrand Russell (Russell, 1905). Russell notes that in Frege's theory, a phrase such as "the present Queen of England" denotes an actual woman. It would seem, by parity of form, that a phrase like "the present King of France" also is about, i.e., denotes, an actual individual. However, as an actual King of France does not exist, this sentence does not denote – it merely expresses a sense. Only when sentences are false or nonsense, do we talk about senses, it seems. Russell then presents a rival theory of denotation which, by means of a clever paraphrasing technique, does not resort to Fregean senses. However, it faces other difficulties. For instance, Russell would have to accept that

Ponce de Leon sought the fountain of youth

is either false or nonsense because there did not exist an actual fountain of youth.

There are other semantical theories and notions similar to Frege's. The notion of connotation is similar to sense. Rudolf Carnap (Carnap, 1947) has presented a semantical method in which he distinguishes the notions extension and intension. These are closely related to Frege's notions, in fact, in direct contexts extension and denotation are the same, and so are intension and sense; only in indirect contexts does Carnap distinguish his notions from Frege's. Common to these theories is the distinction between two semantical notions, one more general than the other.

Extensionality and Intensionality in Formal Settings

So far, the investigations have been restricted to natural language; in the following they will be generalized to logic. The underlying idea is to formalize the condition for extensionality (substitutivity of co-denotational expressions). This will allow us to determine whether a logic is extensional or intensional.

In propositional logic, the formula

$$(1) (P \leftrightarrow Q) \rightarrow (R \leftrightarrow R[Q/P])$$

is valid (logically true) for all formulas P , Q , and R , where $R[Q/P]$ is the result of substituting zero or more occurrences of P with Q in R . This result says that equivalent (co-denotational) formulas can be substituted with preservation of truth, hence, that propositional logic is extensional.

In contrast, modal logic is intensional because (1) is not valid in modal logic. We have the following counterexample comprising the necessity operator:

$$(p \leftrightarrow q) \rightarrow (\Box p \leftrightarrow \Box p[q/p]),$$

that is,

$$(p \leftrightarrow q) \rightarrow (\Box p \leftrightarrow \Box q)$$

is not valid, where p and q are atomic propositions. Thus, in modal logic we cannot substitute equivalent formulas, but this is precisely what we want because modal logic formalizes the intensional notion of modality. The example shows that \Box creates an indirect context.

Presenting a general, formal definition of when a logic is extensional is no trivial task. One reason is that the notion of logic is general indeed, meaning there are several, non-equivalent formalizations of the condition for extensionality. Consider first-order predicate logic, which is commonly said to be extensional. If we accept open formulas (formulas in which variable occurrences are not bound, such as $F(x)$), formula (1) is not valid. But this means that we would have to say that predicate logic is intensional. However, (1) is not the only formalization of the condition for extensionality for predicate logic. As an alternative formalization, we have:

$$(2) \text{ If } P \leftrightarrow Q \text{ is valid then } R \leftrightarrow R[Q/P] \text{ is valid.}$$

The difference between this formalization and (1) is that (2) is formulated in metalanguage. As (2) holds, it says that predicate logic is extensional in terms of metalogical formalization. Unfortunately, we cannot adopt (2) as a general formalization of extensionality because modal logic also satisfies (2), meaning it would become extensional too. A possible solution is to discard open formulas in predicate logic and accept (1) as a formalization of extensionality. However, other solutions might be preferred. It has been suggested, e.g., by Ruth Barcan Marcus (Marcus, 1960), that there are several principles (definitions) of extensionality, and hence also several principles of intensionality. This reveals subtleties in the distinction between extensionality and intensionality.

Because of their imprecise nature, intensional notions such as sense have been deemed opaque. However, the last 50 years of developments in non-classical logic, in particular the development of possible-world semantics by such people as Saul Kripke and Richard Montague (1970), have shown that a significant part of intensional notions can be formulated in precise (mathematical) settings. See Gamut (Gamut, 1991) for an introduction and Fitting and Mendelsohn (Fitting and Mendelsohn, 1998) for newer developments in possible-world semantics.

See also: Modal Logic; Montague Semantics; Sense and Reference: Philosophical Aspects.

Bibliography

- Carnap R (1947). *Meaning and necessity*. Chicago: University of Chicago Press.
- Frege G (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50. Reprinted as ‘On sense and reference.’ In Geach P & Black M (eds.) (1984). *Translations from the philosophical writings of Gottlob Frege*. Oxford: Blackwell. 56–78.

- Fitting M & Mendelsohn R L (1998). *First-order modal logic*. Dordrecht: Kluwer Academic Publishers.
- Gamut L T F (1991). *Logic, language and meaning* (vol. 2). Intensional logic and logical grammar. Chicago: Chicago University Press.
- Marcus R B (1960). ‘Extensionality.’ *Mind, New Series* 69, 55–62.
- Montague R (1970). ‘Universal grammar.’ In Thomason R H (ed.) (1974). *Formal philosophy – selected papers of Richard Montague*. New Haven and London: Yale University Press. 222–246.
- Russell B (1905). ‘On denoting.’ *Mind New Series* 14, 479–493.

Externalism about Content

J Brown, Bristol University, Bristol, UK

© 2006 Elsevier Ltd. All rights reserved.

The Thesis of Externalism

Externalism is a thesis in the philosophy of mind about thought content, i.e., what is believed, desired, hoped, feared, etc. We can distinguish externalism and its opposite, internalism, by contrasting relational properties such as the property of being taller than the Eiffel Tower, and intrinsic properties, such as the property of being made of iron. Externalism treats the property of having a certain thought content as a relational property, whereas internalism treats it as an intrinsic property.

As an example, let Oscar and Twin Oscar share all their intrinsic properties: they are molecule-for-molecule physical duplicates; they share all the same dispositions to behave, where behavior is construed as bodily movement non-intentionally described; and they share all the same ‘narrow’ functional properties, i.e., functional properties defined in terms of causal relations within the body. Externalism claims that Oscar and Twin Oscar may have different thought contents if they are in different environments. Internalism claims that Oscar and Twin Oscar have the same thought contents. Thus we may understand the dispute between externalism and internalism in terms of supervenience. Internalism claims, while externalism denies, that thought content supervenes on the properties that duplicates like Oscar and Twin Oscar share.

Externalism has important consequences for the philosophical debate about the relation of mind and body. Externalism is incompatible with the behaviorist claim that talk of mental states can be reduced to talk of dispositions to behave, where behavior is

construed as bodily movement non-intentionally described. Further, externalism is incompatible with the reduction of types of mental state either to types of brain state, or to narrow functional types. Externalism rejects the Cartesian supposition that a normally embodied person and an always envatted brain (i.e., one not connected to a real body in a real world) could have all the same mental states while being in radically different environments. (Notice that externalism and internalism should not be understood as disputing the uncontroversial truth that a subject’s environment causally affects what she thinks. For instance, an internalist who claims that types of mental state are identical to types of brain state may agree that a subject’s environment causally affects what types of brain state he is in, and so what types of mental state he is in.)

Arguments for and against Externalism

Natural Kind Terms

Putnam’s classic twin earth argument for externalism asks us to imagine that there is a far away planet, Twin Earth, that is a physical duplicate of Earth and that contains duplicates of all Earthians. The only difference between the planets is that whereas, on Earth, the stuff that flows in rivers and falls as rain has chemical composition H_2O , on Twin Earth, the stuff that flows in rivers and falls as rain is an identical seeming stuff that has a different chemical composition, abbreviated XYZ. Consider an Earthian, Oscar, in 1750 when the chemical composition of water was not yet known. Putnam claims that when Oscar sincerely says, “Water is wet,” the term ‘water’ and the concept it expresses refers to all and only H_2O , so that the truth value of Oscar’s utterance and thought turns on the properties of

H₂O. By contrast, Putnam claims that when Oscar's Twin Earthian twin, Twin Oscar, makes the same utterance, the term 'water' and the concept it expresses refers to all and only XYZ, so that the truth value of Twin Oscar's utterance and thought turns on the properties of XYZ. Given that it's an assumption of the thought experiment that Oscar and Twin Oscar are physical duplicates, the fact that their utterances and thoughts are about different substances and have different truth conditions seems to support the externalist claim that a subject's thought contents are partly individuated by the sorts of substances present in his environment.

As initially set-up, the thought experiment suffers from the fact that since humans are largely made out of water, Oscar and Twin Oscar cannot be physical duplicates. However, the thought experiment can easily be reformulated to overcome this objection, by using a term for a 'natural kind' that is not part of the physical constitution of humans. The thought experiment exploits the idea that we intend some of our terms to refer to natural kinds that are naturally demarcated by their fundamental properties, and it is the job of science to discover those properties. On this view, an item could appear superficially just like instances of a certain natural kind and yet not be of that natural kind if it lacks the relevant fundamental properties (e.g., fool's gold). Further, an item could be of a certain natural kind even without possessing its characteristic appearance if it has the fundamental properties definitive of that kind (e.g., an albino tiger). If we intend 'water' to refer to such a natural kind, then XYZ is not included in the reference of our term for, although it looks and behaves like H₂O, XYZ is stipulated to differ in its fundamental chemistry.

One could object to Putnam's Twin Earth argument by rejecting some of the assumptions on which it lies: for example, that there are natural kinds, that some of our non-technical terms aim to refer to such kinds, and that if a substance like XYZ were discovered it would be incorrect to regard it as of the same natural kind as H₂O. Even if all these assumptions are granted, one could further object to the argument that one can accommodate the idea that Oscar's and Twin Oscar's thoughts refer to different substances without granting the externalist claim that their thought contents are partly individuated by the substances in their respective environments. Suppose that each would explicate the meaning of 'water' as "the watery stuff around here." On this explication, since the watery stuff in Oscar's environment is H₂O, in his mouth, 'water' would refer to H₂O. By contrast, since the watery stuff in Twin Oscar's

environment is XYZ, 'water' in Twin Oscar's mouth would refer to XYZ. In this way, someone could hope to explain how Oscar's and Twin Oscar's utterances and thoughts refer to different substances within a view that treats their reference as being determined by a common rule, or 'character' in Kaplan's sense. Such a 'two-dimensionalist' view could arguably be used to support an internalist account of Putnam's Twin Earth story.

Indexicals

A second argument for externalism concerns reference to particular objects, rather than natural kinds. Suppose that a subject sees a blue lamp and says, "That lamp is broken." His utterance and the thought it expresses refer to the lamp he sees, and their truth value turns on the properties of that object. An internalist could suggest that the content of the demonstrative expression and the corresponding thought constituent is given by a certain condition such that the demonstrative expression and thought constituent refer to an object if and only if it meets that condition, say the condition of being the unique blue lamp (on some interpretations, this condition is an example of a Fregean 'sense'). On this view, in a counterfactual condition in which the subject is in the same intrinsic state but looking at a duplicate lamp he would have the same thought content. However, there are a number of objections to the suggested internalist understanding of demonstrative reference. First, it is hard to specify a condition that both uniquely picks out the intuitive referent and that is plausibly available to any subject capable of making or understanding demonstrative reference. Second, understanding the demonstrative expression arguably only requires latching onto the right object, regardless of how one conceives of it. As a result of these difficulties, externalists suggest that the object referred to partly specifies the content of the demonstrative expression and thought constituent, so that if the subject had instead been looking at a duplicate lamp the content of his thought and utterance would have been different. This alternative view supports the externalist claim that a subject's thoughts are partly individuated by the objects in his environment. An internalist could try to avoid this conclusion by exploiting a two-dimensionalist framework in a way analogous to the potential two-dimensionalist reply to Putnam's argument.

Burge and Linguistic Practice

A third argument for externalism, forwarded by Burge, aims to show that a subject's thought contents are partly individuated by the linguistic practices of

his community. Burge's argument concerns a subject, Alf, an ordinary member of the English-speaking community and arthritis sufferer who complains to his doctor, "I fear my arthritis has spread to my thigh." This attitude indicates that Alf incompletely understands arthritis for, by definition, 'arthritis' applies only to problems of the joints. Despite this, Burge argues that Alf has the concept arthritis and that, by his utterance, he expresses the fear that his arthritis has spread to his thigh. Burge supports this interpretation by saying that it would be natural to report Alf's thoughts in this way, despite his incomplete understanding.

Now consider a duplicate of Alf, Twin Alf, who has been brought up in a counterfactual linguistic community in which 'arthritis' is instead defined to apply to rheumatoid ailments of the joints and thighs. In the counterfactual situation, the term 'arthritis' does not express the English concept arthritis. Like Alf, Twin Alf complains to his doctor, "I fear that my arthritis has spread to my thigh." Burge argues that, in the counterfactual situation, Twin Alf lacks the concept arthritis and thus expresses a different thought by his utterance than Alf does. This claim is supported both by the way Twin Alf would explicate 'arthritis' and by how experts in his community would explicate the term. Unlike Putnam's Twin Earth argument, Burge's argument can use a variety of terms and needn't use natural kind terms.

Burge's argument relies on the idea that one may have a concept that one incompletely understands. So, an internalist might object to Burge's argument by rejecting this assumption. In reply, an externalist may point out that the notion that a subject may incompletely understand a concept she possesses is supported by our ordinary practice of attitude ascription in which we routinely ascribe concepts to subjects even despite their incomplete understanding of the relevant term, and regard them as sharing beliefs with others who fully understand the term. Further, it seems supported by Alf's own understanding of the situation. Suppose that the doctor reassures Alf, saying that, by definition, arthritis cannot occur in thighs. It seems likely that Alf would defer to the doctor about the application of 'arthritis,' and regard his earlier fear as false. But, this response suggests that it is the public concept arthritis that figures in his belief, not some idiosyncratic concept defined by his own views.

Accounts of Content

As well as the particular arguments for externalism detailed above, a more general motivation arises from

"naturalistic accounts of content." Such accounts start from the thought that content is not a fundamental feature of the world but can be accounted for in terms of other non-intentional features of the world. On the "causal account," what it is for a state of a system to have a certain content is for it to bear certain kinds of causal relations to other items. On the "teleofunctional account," a state of a system has a certain content if and only if it has a certain evolutionary history. On both of these leading naturalistic accounts, what it is for a state of a system to have a certain content involves its relational properties to items outside that system.

Externalism's Consequences

In addition to its consequences for the mind-body debate (detailed above), externalism has been claimed to have controversial consequences in epistemology and philosophy of mind, in particular for knowledge of one's own mind, knowledge of the world, and the explanation of action. If one's thoughts are partly individuated by the environment, it seems hard for one to know what one thinks just by introspection and without first investigating the environment. Further, if what one thinks depends on the kind of environment one is in, then it may seem that were one to know what one thinks introspectively one could use that knowledge to gain knowledge of the environment without first investigating it! Externalism may also seem to conflict with the plausible intuition that what one thinks causally explains one's actions, given the metaphysical assumption that causal powers depend on intrinsic and not relational properties.

See also: Natural Kind Terms; Two-Dimensional Semantics.

Bibliography

- Bach K (1987). *Thought and reference*. Oxford: Oxford University Press.
- Burge T (1979). 'Individualism and the mental.' *MidWest Studies in Philosophy* 4, 73–121.
- Crane T (1981). 'All the difference in the world.' *Philosophical Quarterly* 41, 1–25.
- Dupré J (1981). 'Natural kinds and biological taxa.' *Philosophical Review* XC, 66–90.
- Evans G (1982). *The varieties of reference*. Oxford: Oxford University Press, Clarendon Press.
- Fodor J (1988). *Psychosemantics*. Cambridge, MA: MIT Press.
- McCulloch G (1989). *The game of the name*. Oxford: Clarendon Press.

-
- Putnam H (1975). 'The meaning of "meaning."' In Gunderson K (ed.) *Minnesota studies in the philosophy of science 2: Language, mind, and knowledge*. Minneapolis: University of Minnesota Press. Reprinted in *Mind, language and reality*, H. Putnam. Cambridge: Cambridge University Press. 215–271.
- Segal G (2000). *A slim book about narrow content*. Cambridge, MA: MIT.
-

This page intentionally left blank

F

Fictional Discourse: Philosophical Aspects

D Davies, McGill University, Montreal, Canada

© 2006 Elsevier Ltd. All rights reserved.

Fictional discourse is among the most superficially unproblematic yet philosophically puzzling of linguistic phenomena. We usually have little difficulty distinguishing fictional from nonfictional writings. We easily make sense of what is set out within the pages of fictions, and readily agree that Pegasus is a winged horse and not a one-legged donkey. Nor are we unreflectively perplexed by the feelings aroused in us by fictions. But each of these issues – the nature of fictional discourse, its meaningfulness, our ability to determine what is true in a fiction, and our capacity to be moved by it – has engendered heated philosophical debate that shows no sign of abating. While some of these issues concern only philosophers of art and literature, others bear on wider issues in ontology and the philosophy of language.

The Nature of Fictional Discourse

As will become apparent, a number of significantly different linguistic phenomena might be classified as forms of fictional discourse, but the core phenomenon is the use of language in fictional narratives such as the Sherlock Holmes stories. The fictionality of such narratives cannot be simply a matter of whether they portray actual happenings and actual agents, for works of fiction seem to contain numerous truths about the real world – e.g., true statements about the geography of Victorian London in Dickens's *Bleak house*. Furthermore, some paradigmatic works of nonfiction, such as medieval texts on alchemy, are largely false. Nor can the style of narration or the use of 'literary' techniques serve as a criterion of fictionality; there are non-fictional works that make free use of the narrative structures and stylistic devices typical of fiction (e.g. Mailer's *The executioner's song*), and celebrated works of fiction that adopt the style of nonfictional genres of writing (e.g. the 'academic' style of Nabokov's *Pale fire*).

The fictionality of a narrative may therefore depend upon **how it functions**, or **how it was designed to function** – either the function conferred upon it by its users (e.g., Walton, 1990), or its author's intention that it function, or be used, in a certain way. Philosophers attracted by the latter analysis have drawn upon the more general notion of a speech act, that is, an action performed by using language (verbal or nonverbal) in a certain way (Searle, 1975). The paradigm speech act is assertion, and Searle has maintained that fiction results when an agent **pretends** to assert what is stated by the sentences that make up a narrative. Given that there seem to be pretended assertions that do not result in fictions – e.g. a satirist's mimicking of a politician, Currie (1990) proposed that fiction results from the genuine performance of a speech act, but one that differs from assertion in that, whereas the latter requires that the speaker intend her audience to **believe** what she states, the author of a fictive utterance intends that her audience **make believe** what is narrated (see also Lamarque, 1996). To understand a work of fiction, then, is to grasp what it is that, as reader, one is intended to make believe.

Truth in a Fiction

If this is on the right track, what reading strategies should one adopt to determine what one is required to make believe to count as understanding what is 'true in a story'? Not everything explicitly affirmed by the narrator of a story can be true in the story, for there are deceived, or deceiving, or ironic narrators. Nor should we restrict what is true in a story to what is explicitly affirmed by a trusted narrator, for sometimes, as in ordinary discourse, we take a narrator to be communicating more by her words than is strictly stated. We also assume, in our reading, that many things are true in a story although they are neither explicit nor implicit in the trustworthy affirmations of the narrator. We must therefore explain how we determine the unspecified background of general propositions that are nonexplicitly true in a given story. One strategy, here, appeals to what philosophers term

“possible worlds” – roughly speaking, alternative ways the actual world might have been (Lewis, 1983): what is true in a (fictional) story *N* is decided by reference to possible worlds in which those things that are explicitly or implicitly true in the text of *N* are actually true. The task, then, is to pick out a particular group of such possible worlds as the ones that determine what else is true in the story.

A general problem for any such account of fictional truth, however, is that it cannot deal with stories in which **inconsistent** truths obtain – for example, certain sorts of time-travel stories – since possible worlds must themselves be **consistent** in order to be possible. To resolve this problem, Currie has proposed an analysis not in terms of possible worlds in which a given set of beliefs is true, but, rather, in terms of belief-sets themselves. He maintained that, in reading fictional narratives, we make believe that we are being informed about the events in the story by a reliable source, the so-called fictional author of the story. It is by reference to the beliefs attributable to the fictional author (the f.a.) – a completely reliable source of information about the narrated events – that the reader determines what is true in the story. The reader is supposedly guided here by the text itself, which provides evidence as to the f.a.’s character and psychological idiosyncracies, and by background assumptions – based on knowledge of the real author – as to the historical situatedness of the f.a. But it is open to question whether these resources are adequate to the task (see Davies, 1996).

Semantics of Fictional Discourse

The suggested strategies are supposed to explain how, given an instance of fictional discourse, we can understand what is true in the narrative presented through that discourse. Even if the author does not assert the sentences in such a discourse, it is assumed that the sentences themselves provide us with the content for our acts of make-believe. The use of language in fictional discourse is thus assumed to differ pragmatically rather than semantically from its use in nonfictional discourse (Gale, 1971). This is unproblematic when fictional discourse predicates nonactual properties of actual things – when, for example, Chancery is represented by Dickens as being shrouded in an impenetrable fog. But problems arise when a sentence *S* in fictional discourse contains what appears to be a proper name *N* (e.g. ‘John Jarndyce,’ as it occurs in *Bleak house*), which is not plausibly taken as denoting an actual person, place, or event. Suppose ‘fictional names,’ as Currie terms them, are genuine proper names, and that the same semantic principles apply to fictive and nonfictive

discourse. Then, given the widely accepted ‘direct reference’ theory of proper names (Kripke, 1972), where a proper name contributes to the meaning of a sentence only the individual it denotes, *S* is meaningful only if *N* is a genuinely denoting term. If, by definition, no actual person, place, or event exists that is denoted by a fictional name – if such names are empty – how can we explain our seeming ability to understand sentences containing fictional names, and to hold some of them expressive of what is ‘true in the story’?

Something, it seems, has to give. That is:

- a. fictional names do genuinely denote, or
- b. *S*’s logical form is not as it appears to be, and fictional names do not function as proper names, or
- c. our ability to make sense of *S* is compatible with *S*’s being meaningless.

Option (a) is most famously associated with Meinong (1960), who held that there are nonexistent objects corresponding to unactualised sets of properties. Contemporary Meinongians, such as Parsons (1980) and Zalta (1983), have asserted that fictional characters are nonexistent objects of this kind, and that fictional names denote fictional characters so construed. An alternative strategy is to take fictional names to denote nonactual but possible objects that exist in other possible worlds (Plantinga, 1974; Lewis, 1983). Less ontologically profligate is the contention that fictional characters exist as ‘abstract artifacts’ (Thomasson, 1999) or as theoretical entities required to make sense of our critical practice (Van Inwagen, 1977), where such entities depend for their being on the activities of authors or critics. Those who choose option (a) must explain how fictional names are able to refer to fictional objects, as abstract entities, if names secure their reference through some kind of causal/historical link with their denotata, and how fictional characters are individuated. Arguably, Meinongians are able neither to meet these challenges nor to validate our intuition that authors create the characters in their works (Thomasson, 1999), whereas, so it is claimed, these difficulties can be surmounted if fictional characters are taken to be abstract artifacts.

Option (b) requires that sentences containing fictional names be paraphrased such that they no longer contain such denoting expressions but are meaningful and possess appropriate truth values. Russell (1990), for example, analyzed sentences such as ‘Pegasus flies’ as quantified sentences affirming the existence of something that is a winged horse and flies: fictional names, like all standard names for Russell, are disguised definite descriptions. But this strategy makes sentences containing fictional names come out false

unless by chance there is an existent fitting the relevant description. One response, here, is to take sentences occurring in fictional discourse to have a suppressed 'in the story' operator, while retaining the idea that fictional names, in such contexts, are disguised descriptions (Currie, 1990).

Option (c) grants that fictional names are genuine names and that they are empty, and therefore that sentences containing fictional names are meaningless. We are, however, able to understand fictions and determine what is true in a story, since, in engaging with a fiction, we make believe that the contained sentences are meaningful and that they have the relevant truth values (Walton, 1990).

A problem that is especially acute for those who choose option (b) is accounting for occurrences of fictional names in other kinds of linguistic contexts where it is difficult to deny our commitment to the reality of fictional characters. Discourse **about** fiction includes such utterances as 'N is a fictional character,' 'Tolstoy created N,' and 'N is a better detective than M.' Paraphrases of such sentences offered by proponents of both (b) and (c) seem implausible. However, Meinongians also have to tell a different story about the occurrence of fictional names in such contexts. Zalta (1983), for example, who has held that fictional characters, as nonexistent objects, 'encode' but do not exemplify the properties ascribed to them in the story, maintained that, in discourse about fiction, the characters do exemplify the predicated properties. Van Inwagen (1977) and Thomasson (1999) preserved the surface grammar of such occurrences of fictional names, but at the price of denying that fictional characters literally possess any of the properties predicated of them in the story: for Van Inwagen, such properties are ascribed but not possessed.

Responding to Fictions

Philosophers have also puzzled over the capacity of fictions to elicit fear or other emotions in readers. The puzzle is especially acute given the recent rise to dominance of cognitivist theories of the emotions. According to the latter, my being in an emotional state like fear requires more than that I have a certain feeling or am in a certain affective state. My fearing *x* also requires that I believe that *x* endangers me or someone close to me. Given these requirements, it seems impossible for me to experience genuine fear (or, *mutatis mutandis*, pity) concerning what is narrated in what I take to be a fiction, because I do not believe that anyone is really in danger when I imaginatively engage with such a fiction.

One strategy here is to argue that our affective responses to fictions do have the cognitive foundations

required if they are to be genuine emotions. Perhaps, for example, we believe that real people are endangered or suffering **in the same way** that the fictional characters are represented as doing, and this leads us to feel genuine emotions for these real people. More plausible are analyses that take as central the belief that it is **make-believe**, in the fiction, that I stand in the relation to the fictional characters or situations required if I am to feel genuine emotions for them. According to Walton, such beliefs about the story elicit in us quasi-emotions, which resemble genuine emotions in their physiological and psychological features and which enter into our imaginative engagement with the fiction by encouraging us to **make believe** that we feel the corresponding real emotions (see also Currie, 1990, chapter 5). Neill (1993), however, contended that affective responses elicited in us through our beliefs as to what is fictionally true are often instances of genuine emotions. A third alternative is the so-called 'thought theory,' which holds that such responses (whether or not they are genuine emotions) are engendered not by beliefs about what is the case, actually or fictionally, but by thoughts (Carroll, 1990; Lamarque, 1996) or imaginings (Novitz, 1987) as to what might be the case. (See Hjort and Laver, 1997, for further articles.)

See also: Assertion; Direct Reference; Empty Names; Possible Worlds; Philosophical Theories; Proper Names; Philosophical Aspects.

Bibliography

- Carroll N (1990). *The philosophy of horror*. New York: Routledge.
- Currie G (1990). *The nature of fiction*. Cambridge: Cambridge University Press.
- Davies D (1996). 'Fictional truth and fictional authors.' *British Journal of Aesthetics* 36, 43–55.
- Gale R (1971). 'The Fictive Use of Language.' *Philosophy* 46, 324–30.
- Hjort M & Laver S (eds.) (1997). *Emotion and the arts*. Oxford: Oxford University Press. 37–92.
- Kripke S (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Lamarque P (1996). *Fictional points of view*. Ithaca: Cornell University Press.
- Lewis D (1983). 'Truth in fiction.' In *Philosophical Papers, Volume 1*. New York: Oxford University Press. 261–275.
- Meinong A (1960). 'Theory of Objects.' In Chisholm R (ed.) *Realism and the background of phenomenology*. Glencoe, IL: Free Press.
- Neill A (1993). 'Fiction and the Emotions.' *American Philosophical Quarterly* 30, 1–13.
- Novitz D (1987). *Knowledge, fiction, and imagination*. Philadelphia: Temple University Press.

- Parsons T (1980). *Nonexistent objects*. New Haven: Yale University Press.
- Plantinga A (1974). *The nature of necessity*. Oxford: Clarendon Press.
- Russell B (1990). 'On Denoting.' Reprinted in Martinich A P (ed.). *The philosophy of language* (2nd edn.). New York: Oxford University Press. 203–211.
- Searle J (1975). 'The logical status of fictional discourse.' *New Literary History* 6, 319–332.

- Thomasson A (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.
- Van Inwagen P (1977). 'Creatures of Fiction.' *American Philosophical Quarterly* 14, 299–308.
- Walton K (1990). *Mimesis as make-believe*. Cambridge, MA: Harvard University Press.
- Zalta E (1983). *Abstract objects*. The Netherlands: Reidel.

Figurative Language: Semiotics

F Nuessel, University of Louisville, Louisville, KY, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The term 'figurative language' can be misleading because it has at least two meanings. For many years, texts on literary criticism have used this expression to mean a literary adornment or a stylistic device (Wellek and Warren, 1956: 23, 157). Lakoff and Johnson (1980: 3) noted this conventional usage in their influential work on metaphor. In general, the expression 'figurative language' was restricted to special usage in poetry or rhetorical language. Traditional scholars maintain a strict dichotomy between figurative language and ordinary or literal language. This conventional aesthetic sense of figurative language no longer reflects current usage. Today, the term 'metaphor' has replaced 'figurative language' with the special sense of a cognitive device used to explain how people categorize reality and store abstractions of that physical existence in their brain.

In this article, we will discuss the following issues:

1. Theories of metaphor
2. A selected overview of theoreticians of metaphor and their theories
3. The literal-figurative dichotomy
4. The role of metaphor as a cognitive mechanism
5. The use of metaphor as a knowledge generator.

The literature about metaphor is voluminous, as exemplified selectively in three extensive bibliographies on this topic (Noppen and Hols, 1990; Noppen *et al.*, 1985; Shibbes, 1971). Moreover, important journals devoted to metaphor, such as *Metaphor and Symbol* (formerly, *Metaphor and Symbolic Activity*), *Poetics*, and *Cognitive Linguistics* reflect this burgeoning interest.

The explosion of literature on metaphor reflects some basic shifts in linguistic theory during the 20th

century. In particular, there have been three philosophical revolutions in linguistic theory. In the first part of the 20th century, the North American linguist Leonard Bloomfield (1887–1949) espoused what he called a mentalist approach to the study of language in his book *Introduction to the study of language* (1914). Subsequently, Bloomfield would undergo a philosophical transformation while at Ohio State University under the influence of the behavioral psychologist Albert Weiss (1879–1931). This theoretical conversion appeared in Bloomfield's classic volume *Language* (1933). This behaviorist, or empiricist, view of language received its maximum articulation in B. F. Skinner's (1904–1990) *Verbal behavior* (1957). Its stunning repudiation by Noam Chomsky (b. 1928) in his review in the journal *Language* in the same year represents the beginning of the Cartesian, or rationalist, approach to linguistic theory that continues to be vital today. Neo-empiricism, however, reasserted itself in the influential book *Metaphors we live by* (1980) by George Lakoff (b. 1941) and Mark Johnson (b. 1949).

Traditional discussion of figurative language features a detailed typology of figures of speech. It was a standard practice for those who studied Greek and Latin to receive training in the tropes (derived from Greek for 'manner') of literary language. These include metaphor, simile, personification, and others. This classical view of figurative language as the adornment of a basic literal language influenced literary research and analysis. Wellek and Warren's *Theory of literature* (1956) embodied this approach, and it was highly influential in university literature departments.

Conventional courses in literature often have a component on rhetorical or figurative language. These discussions view such language as stylistic in nature and designed to arouse the reader's interest through these embellishments. The following is a selected enumeration of some conventional tropes

(cf. Danesi and Perron, 1999: 162–164 for a more extensive list).

Theories of Metaphor

Most theories of metaphor can be reduced to two basic ones with some variation. In this section, we will examine these theories and some of their most prominent proponents. First, there is the comparison theory, which may be traced back to the Greek philosopher Aristotle (384–322 B.C.). In this model, metaphor has the form *A is B*, and it suggests an analogy between two items or notions. In this view, a metaphor is a compressed simile. This is the traditional approach, and it includes the resemblance of objects or notions; its usage involves style or economy of statement. A variant of the comparison theory is the substitution theory. The substitution theory means that an inappropriate (figurative) term stands for a proper (literal) one. In this approach, the speaker uses an expression that means, or refers to, something else, thereby requiring the listener to decode the statement. An example is *Martin Luther King is a lion*. *Lion* thus substitutes for the meaning that Martin Luther King is brave and fearless. The essence of the comparison model or the substitution version is based on the premise that there is a basic literal language that is precise and neutral. This language stands in stark contrast to the poetic language generally associated with literature and rhetorical flourishes.

The second major theory is the interactionist theory, which, with its various manifestations and modifications in the 20th century, is the prevailing one today. It contrasts with the comparison view of metaphor. In this approach, there is an interaction of concepts at sentence-level and not word-level, i.e., there is a conceptual association of the elements that constitute a metaphor, and this association gives rise to new meaning as we shall see.

Selected Review of Theories of Metaphor

In this section, we will examine selectively prominent theoreticians of metaphor. These include Aristotle, Giambattista Vico, Groupe μ , I. A. Richards, Max Black, Paul Ricoeur, George Lakoff, and Mark Johnson.

Aristotle (384–322 B.C.)

Aristotle is credited with introducing the word ‘metaphor’ in two of his major treatises (*Rhetoric* and *Poetics* [1997]). The etymology of the word metaphor is from the Greek ‘to carry beyond’ or ‘to transfer.’ This meaning refers to an implied comparison or transference of meaning between two objects formulated as ‘A is B’ or ‘A implies B.’

Aristotle’s view of metaphor is a literalist one, i.e., metaphors stand in contrast to ordinary language. Moreover, his view of metaphor is now known as the comparison theory of metaphor, i.e., ‘A is B’ or ‘A implies B.’ Aristotle’s influential statements about language and rhetoric in *Rhetoric* and *Poetics* continue to be cited after more than two millennia.

The Greek philosopher, and student of Plato (ca. 428–347 B.C.), considered metaphor in *Poetics* (Aristotle, 1997: 150–156). In a general discussion of grammar and etymology, Aristotle provided selected observations on metaphor. In particular, Aristotle stated that “metaphor is the transference of a word of another significance either from genus to species, or from species to genus, or from species to species or by analogy or proportion” (Aristotle, 1997: 150). Aristotle (1997: 151) stated that there are four types of metaphor. His consideration of the first three was terse, while his commentary on the fourth was a bit more expansive.

Aristotle’s first type, that of genus to species, is the expression “my ship stands there” (Aristotle, 1997: 151). In this type of metaphor, which is, in fact, a

Table 1 Some examples of traditional figurative language

Type of figurative language:	Definition:
Conceit	An analogy between two completely distinct notions, e.g., <i>war is peace</i> .
Euphemism	Use of a pleasant word for one that is vulgar or obscene, e.g., <i>powder room for toilet</i> .
Hyperbole	Extreme exaggeration, e.g., <i>he ate half a side of beef</i> .
Irony	Sarcastic wit to mean the opposite of what is meant, e.g., <i>I love that dress!</i> (meaning, <i>I hate that dress!</i>).
Litotes	Understatement to magnify the significance of a statement, e.g., <i>Verdi was not an unimportant composer</i> .
Metonymy	A common feature is used to designate the whole entity, e.g., <i>the pen is mightier than the sword</i> .
Onomatopoeia	The use of words to describe the sounds that they refer to, e.g., <i>the clanging of the bells</i> .
Oxymoron	An observation that is apparently contradictory, e.g., <i>he is a wise fool</i> .
Paradox	An apparently contradictory statement, e.g., <i>dumb as a fox</i> .
Personification	The representation of a concept or an object as living entities, e.g., <i>Death came through the door</i> .
Rhetorical question	A question that is not intended to be answered, e.g., <i>do you understand me?</i>
Simile	An overt comparison with words such as <i>like</i> or <i>as</i> , e.g., <i>he floats like a butterfly</i> .
Synecdoche	The part stands for the whole, e.g., <i>the law (= the police) broke down the door</i> .

synecdoche, 'stand' is the genus and 'lying at anchor' is the species.

The second type (species to genus) offered by Aristotle (1997: 151) is "verily ten thousand noble deeds hath Odysseus wrought," in which case 'ten thousand' signifies 'many.' Again, this is a synecdoche, and some critics have labeled it 'clumsy' because the choice of 'ten thousand' to mean 'many,' even though it may have been a commonly accepted expression in Aristotle's time, is definitely not novel.

The third metaphor type, and the exemplar for species to species is "with the blade of bronze drawing out the life" (1997: 151) and "cutting with the unwearied bronze" (1997: 151). In these cases, 'drawing out' means 'cutting,' while 'cutting' means 'removing something.' Both mean to 'take away.' Umberto Eco (b. 1932) argued that this third type of metaphor was the quintessential one on which subsequent theories base their own views of the nature of metaphor (1984: 92). Aristotle's third type of metaphoric exemplar perhaps best illustrates 20th-century views of metaphor as a tripartite concept, e.g., I. A. Richards's (about whom more will be said below) notion of a metaphor as consisting of a tenor (the metaphORIZED term), a vehicle (the metaphORIZING term), and a ground (the meaning created by the interaction of the tenor and the vehicle). This type of metaphor has the form (A (B) C), i.e., A is the tenor, C is the vehicle and B is the ground or the intermediary term which provides the meaning of the metaphor supplied by the interpretant, or the person who seeks to make meaning of the comparison. The tenor (often called the topic now) is the A-referent (what is talked about in the metaphor). The vehicle is the C-referent (something concrete or familiar) and the ground is the meaning of the comparison, or the B-referent. In the sentence, *Mary is a witch*, 'witch' is the vehicle, 'Mary' is the tenor or topic, and the meaning created by the connection of these two terms is the ground. Cultural context will assign a specific meaning to the vehicle of a metaphor.

The fourth and last type of metaphor, with Aristotle's overt statement that it is by analogy or proportion, involves a formula with four elements. Aristotle offered two examples. The first is the analogy of the "shield of Dionysus" and the "cup of Ares" (Aristotle, 1997: 151). In this four-part proportional metaphor, the four terms may be delineated in the following way as a subject/instrument relationship: Dionysus : cup :: Ares : shield. This is to say that 'shield of Dionysus' is a metaphor for 'cup,' and 'cup of Ares' is a metaphor for 'shield.' Thus, the cup is to Dionysus as the shield is to Ares. Eco (1984: 95) pointed out that there are similarities and differences in this metaphor. Dionysus and Ares are both gods (the former of joy and

peaceful rites, the latter of war and death). The cup and shield are alike due to their roundness. The hearer must pick out the similarities and discard the differences. If there are too many differences, the metaphor is unsuccessful.

In Aristotle's second example (Aristotle, 1997: 151), "old age [is]... the sunset of life," attributed to Empedocles (5th century, B.C.), there is a proportion or analogy that consists of four terms, namely, A : B :: C : D, i.e., old age : life :: sunset : day. This proportion is different from the above metaphor because it involves the proportion A : B :: C : D. In this metaphor, the proportion may be stated in the following way: A = old age, B = life, C = evening, D = day. In this sense, A is to B as C is to D. Hence this proportional analogy is a comparison of the life cycle to the diurnal cycle.

Thus, in a succinct narration, Aristotle provided us with our initial knowledge of the metaphor, and one to which all future scholars would refer over the millennia. Aristotle's views on metaphor contain three basic notions: (1) there is an exchange of properties between two entities to increase understanding; (2) there is a cognitive dimension to metaphor, albeit suggested rather than overtly stated, and Aristotle does not explain this; and (3) the pair metaphor/metonymy represents the essence of semiosis and human thought.

It is also clear that Aristotle left a legacy that includes a doctrine of signs, when he observed that words are tripartite objects which consist of a form (pronunciation), a referent (what the word alludes to), and a meaning. In one of his statements on metaphor, Aristotle (1997: 153) stated that "... making good metaphors requires the ability to grasp resemblances." In effect, this observation is a comparison model of metaphor. Aristotle's comparison theory has met with criticism over the centuries, though its basic premise remains intact.

It should be noted that the Roman rhetorician Quintilian (ca. 35–100 A.D.) espoused Aristotle's comparison view of metaphor and claimed that it was a substitutive process. This substitution theory argued that in the metaphor 'John is a rat,' the use of 'rat' for 'man' (in the sense of a vile and loathsome creature) was simply a deviation from a pristine literal language. Neither Aristotle nor Quintilian offered a psychological explanation for the pervasive use of metaphor in world languages. That would not take place in any systematic way until the 20th century.

Giambattista Vico (1668–1744)

The Neapolitan empiricist philosopher had many interesting things to say about metaphor, and his observations presage important work in the 20th century,

and it is for this reason that we discuss his work here. In his treatise *The new science* (1968), Vico spoke about figurative language, in particular, metaphor. Reawakened interest in Vico in the 20th century corresponds to a renewed interest in metaphor research and the way in which Vico conceptualized it. Vico differs from his rationalist counterpart, the French philosopher René Descartes (1596–1660), whose views on language are quite distinct. Descartes was a rationalist who believed that people possess an innate linguistic faculty rather than an acquired one.

Vico's views on language consist of a series of basic epistemological principles that propose how we acquire knowledge about the world and how language develops. His stance on language in general, and metaphor in particular, may be stated as axioms about the origination of language and thought. We will now discuss this set of stages or steps in the development of thought and language, and, ultimately, metaphor.

First, thought is corporeal. We acquire our knowledge of the external world through our senses. In this regard, Vico espouses his empiricist view of epistemology by arguing that our knowledge has a sensory basis, i.e., we learn about the world by experiencing it first and foremost through vision, as well hearing, touch, smell, and taste. We then use this information to categorize this knowledge in a systematic way through our contact with the external world. Second, thought is representational. This means that it is imaginative or image-creating. These visual, culturally-determined patterns allow us to access information retained in our memory iconically. The linguist Ann Viberg noted that in the languages of the world, most of the predicates that are synonymous with the verbs of knowing and understanding relate to visual intake. In fact, prior to the development of a systematic language system with a phonology, morphology, syntax, semantics, and lexicon, it is likely that prelinguistic peoples utilized visual signals and signs including primitive pictures for communication as evidenced by the drawings and paintings at various archaeological sites worldwide, e.g., the cave paintings of Altamira, Spain. Third, thought has gestalt (German for configuration)-like properties, i.e., it consists of segments or parts that have a greater significance than its individual constituents. This axiom means that we configure thought through a series of components that build from the concrete to the abstract. Fourth, thought has an ecological basis. For Vico, human language evolved through a series of stages that included gestures that stood for a single article (tool) or concept (fire), and these symbolic gestures developed over time into complex verbal systems. Finally, thought and its verbal manifestation speech is, in fact, metaphorical in nature. For Vico, the process

by which thought and language developed means that language is metaphorical, because we express ourselves through categories based on real-world experience, which we internalize in our brain in meaningful components.

The essence of Vico's cognitive scientific views of metaphor appears in the following citation "...in all languages the greater part of the expressions relating to inanimate things are formed by metaphor from the human body and its parts and from the human senses and passions. Thus, the head for top or beginning; the brow and shoulders of a hill..." (Vico, 1968: 129). This association of external reality with the bodily experiences of that reality is the essence of the new empiricist linguistics which was encapsulated in Lakoff and Johnson's (1980: 59) observation that "...we typically conceptualize the nonphysical *in terms of* the physical..." i.e., the unknown in terms of the known.

I. A. Richards (1893–1979)

Richards introduced some basic notions on the structure of metaphor in his classic treatise *The philosophy of rhetoric* (1936). In the fifth and sixth lectures ('Metaphor' and 'The command of metaphor') of the Mary Flexner Lectures series at Bryn Mawr University, he discussed metaphor in revolutionary terms. In his fifth lecture, Richards (1936: 90) noted that metaphor has traditionally meant "...a grace or ornament or *added* power of language, not its constitutive form." In this groundbreaking paper, Richards would go on to espouse his interactionist theory of metaphor. He argued, in fact, that metaphor is ubiquitous in language, and, by implication there is no literal-figurative speech dichotomy. Richards then introduced new expressions for the three parts of a metaphor, namely, 'tenor,' 'vehicle,' and 'ground.' In a prototypical metaphor, the tenor is the topic or the unknown part in Richard's triadic configuration. The vehicle is the part that comments on the topic, and the ground is the actual meaning created by the interaction of the tenor and the vehicle.

In this theoretical model, Richards builds upon Aristotle's claim that metaphor is the comparison of two apparently dissimilar items (*A is B*) or the substitution of one item for another (*A stands for B*). In Richards' model, however, there is an interaction, i.e., parts of the two entities remain separate and distinct but there is a domain of shared properties in which a new meaning is created. This area of shared properties may be configured as (*A(X) B*). In this model, A is the tenor (topic) or the unknown aspect of the metaphor. B is the vehicle that is a familiar object or concept. X in the equation just given is the ground or meaning produced by the association of the tenor (A)

and the vehicle (B). Thus, the previously mentioned statement *John is a rat* causes the hearer to think or experience one thing in terms of another. In this case, we are invited to draw comparisons between the behavior and traits of a person (*John*) and those of an animal (*rat*). Clearly, a human and a rat are distinct, yet they may share certain traits, e.g., animate, furtive, sneaky, filthy, to name but a few domains of similarity. It is in these areas that new meanings are created. In this specific case, the prototype metaphor or conceptual metaphor (see Lakoff and Johnson below) is PEOPLE ARE ANIMALS (conceptual metaphors are presented in upper case letters by convention and we will follow this format throughout). The specific case here causes the hearer or reader to draw inferences about the shared properties of *John* and a *rat*. To be sure, the similarities in this instance are not pleasant ones. We can thus state that *John* is the tenor or topic, *rat* is the vehicle, and the ground is the meaning of this metaphor, i.e., the conceptualization of John as a vile person. The juxtaposition of the A component (tenor/topic) and the B component (vehicle) is thus responsible for an interaction that provides a new meaning that had not previously existed. Although Richards uses the term ‘interaction’ in his writing on metaphor, this expression would be most closely associated with Max Black (see below) who gave it currency.

As we shall demonstrate, scientific discovery relies heavily on the use of metaphor to create new knowledge about the world by this interactionist strategy. Many scientific breakthroughs have occurred by using this epistemological strategy, e.g., William Harvey’s (1578–1657) use of the machine metaphor to describe the physiological phenomenon of the movement of the valves of the heart.

Richards (1936: 92, 93) also points out that metaphor is “omnipresent” in speech. He notes that even in scientific and technical language, metaphors are ubiquitous. Although Richards did not carry out an empirical analysis of this assertion, Pollio and his associates (1977) did just this, and they discovered that novel and dead metaphoric language permeated speech. Richards thus takes the position that there is no literal-figurative dichotomy, i.e., the claim that there is a basic ‘ordinary’ language that is embellished by the use of colorful figurative language.

In yet another way, Richards is a precursor of the theoretical metaphorology espoused by George Lakoff and his advocates. He states that thought is metaphoric. By this, he means that metaphor constitutes the essence of epistemology. We acquire knowledge about our world and ourselves by converting concrete experience into abstract manipulable units of that experience that allow us to categorize it and refer to it.

Max Black (1909–1988)

Max Black provides another perspective of the interactionist view of metaphor. In his illuminating discussion on the form of metaphor (Black, 1981 [1955]: 77–78), Black enumerated seven basic points about the interaction model of metaphor. First, a metaphor consists of two topics, a primary and a secondary one. Next, these topics are not monolithic entities, but rather they are complex, well-organized groupings of properties. Third, the metaphor affixes to the primary topic a complex of analogous insinuations attributable to the secondary topic. Fourth, in general, these inferences involve commonalities, though, occasionally, there are startling associations. Fifth, metaphors are selective in their implications, i.e., they choose only certain aspects of the secondary topic that apply to the primary one. Sixth, transformations in the lexical meaning belonging to the same family may be metaphoric transfers. Finally, no simple rule for shifts in meaning exists, nor is there a good explanation for why some metaphors are accepted and others are not.

Schroots (1991: 222) best captured the essence of Black’s interactionist view of metaphor in the following observations:

1. Metaphors are concerned with systems of ideas;
2. Such systems are specially constructed;
3. Metaphors are projective in that they allow one field of thought (subject) to organize another, whether by selection, filtering, focusing, or other means; and
4. Models are extended and systematic metaphors.

The celebrated semiotician Charles Peirce (1839–1914) espoused the concept of ‘abduction.’ Simply stated, this is an educated guess. With regard to metaphor, abduction allows the hearer to make a meaningful association between the tenor (now generally called the topic) and the vehicle, i.e., the A and the B components of a metaphor. Abduction would thus allow the hearer to engage in meaningful hypotheses about what the associations between the tenor and the vehicle might be. As noted above, if the associations are tenuous or oblique, then the metaphor will be unsuccessful.

Groupe μ

Groupe μ consists of a cadre of scholars at the University of Liège in Brussels who published *A general rhetoric* (1981) that dealt with figurative language. It predated Lakoff and Johnson’s groundbreaking work on metaphor but offered an excellent overview of a structuralist approach to metaphor research. This group represents the ideology of Ferdinand de

Saussure (1857–1913), Louis Hjelmslev (1899–1965), Emile Benveniste (1902–1976), Algirdas Julien Greimas (1917–1992), Roman Jakobson (1896–1982), and others of the Structuralist school (see more about this linguistic school below).

Paul Ricoeur (1913–2005)

Paul Ricoeur's study (1977) of metaphor is considered to be one of the major statement on metaphor. In this important volume, Ricoeur authored eight essays on metaphor. Ricoeur appropriately began with a consideration of Aristotle's views as the prime mover in metaphor research for more than two millennia. Next, he reviewed the French writer Pierre Fontanier, whose taxonomic approach was inadequate to explain metaphor. In his third study, the author examined metaphor as a sentence-level or a word-level device. In his fourth and fifth studies, Ricoeur investigated structuralist views of metaphor. The sixth study moved to the discourse-level of metaphoric analysis. The seventh paper discussed reference, and the final one dealt with the philosophical underpinnings of metaphor research. What makes this volume valuable is the author's ability to analyze the multiple approaches to metaphor in an objective fashion. Ricoeur brought to our attention that the study of metaphor means coming to grips with the essential nature of language – meaning and semantic deviation.

George Lakoff (b. 1941) and Mark Johnson (b. 1949)

Lakoff and Johnson articulated an elaborate theory of metaphor as a cognitive construct. Their collaborative and individual work has done much to advance metaphor research in the past quarter century. In this section, we shall devote an extended discussion to their contributions to metaphorology.

In *Metaphors we live by* (1980), Lakoff and Johnson discussed the notion of the conceptual metaphor. Their approach to the study of metaphor is now known as Conceptual Metaphor Theory. Thus metaphors such as the previously stated example *John is a rat* (discussed previously) is part of the general conceptual metaphor which equates qualities of people with qualities of animals. Sometimes, these are positive (*Mary is a bunny rabbit*), and in other instances they are negative (*John is a warthog*). The preceding three examples belong to the conceptual metaphor PEOPLE ARE ANIMALS.

Conceptual Metaphor The conceptual metaphor PEOPLE ARE ANIMALS is relatively common. In recent times, the generic term predator, those animals that live by preying on other, weaker animals for

sustenance, has come into frequent use. Predatory animals include, but are not limited to, lions, tigers, and pumas. When people are labeled as predators, this is a subcategory of the PEOPLE ARE ANIMALS conceptual metaphor. This type of metaphor is common in the media, especially when referring to certain types of heinous crimes. This metaphor involves those people who prey upon people who are weaker, e.g., children, older adults, or people with disabilities. The following sentences illustrate this usage.

1. The man *stalked* his victim for days.
2. The outlaws *hunted* their prey surreptitiously.
3. The criminal *pursued* his quarry relentlessly.

The sense of this subcategory of PEOPLE ARE ANIMALS is decidedly negative, because it evokes powerful negative emotions due to its unpleasant associations. Conceptual metaphors thus consist of two parts that are called domains. The first is a source domain (vehicle, or the known, sometimes labeled the B-referent) and a target domain (tenor or topic, or the abstract element, sometimes called the A-referent). In the PEOPLE ARE PREDATORS conceptual metaphor, *people* is the target domain, or the topic itself, while *predator* is the source domain, or the vehicle that delivers the metaphor to the hearer. A metaphor is thus the mapping of one domain into another.

In the Lakoff and Johnson model, metaphor is not a mere deviation from an underlying literal language that functions as a linguistic ornament. Rather, metaphor is now viewed as a cognitive process that allows the mapping of a source to a target domain to enhance comprehension of the unknown in terms of the known. Thus, for Lakoff and Johnson (1980: 5, emphasis in original) “the essence of metaphor is understanding and experiencing one kind of thing in terms of another.”

Other examples of conceptual metaphors include the ARGUMENT IS WAR metaphor, exemplified by such common examples as the following sentences.

ARGUMENT IS WAR

1. They *fought* my claims constantly.
2. I *won* that point.
3. They *beat* us in court.

An additional illustration of conceptual metaphors follows. It should be noted that many of these are rooted in the proverbial language of a culture. In English, ‘time is money’ is simultaneously a proverb and a conceptual metaphor.

TIME IS MONEY

1. We will *buy back* your time with release time.
2. *Budget* your time for each section of the test.

3. *Donate* some of your free time to this cause.

The number of conceptual metaphors in a language is potentially unlimited. Different cultures, however, will focus on and highlight different aspects of these metaphors.

Metonymy A subcategory of metaphor is metonymy. Metonymy refers to the use of one element to refer to another that is closely related. The following examples illustrate this cognitive mechanism.

1. He loves *Cervantes*. (Cervantes = the writings of Cervantes)
2. She's in *opera*. (opera = the profession of operatic performance)
3. *New hair* made John very happy. (new hair = surgical procedure to implant hair).

Synecdoche Synecdoche is a subtype of metonym. It is the conceptual process by which the part stands for the whole, i.e., a segment of an object stands for the entire object. The following examples depict this cognitive device.

1. The *SUV* is destroying the environment. (= the overuse of the this type of vehicle is destroying our limited natural resources)
2. A few good *arms* will help us win this baseball game. (= good pitchers will help the team to win the baseball game)
3. The *brains* in that university are world-class. (= the intelligent people on the faculty).

Marcel Danesi (1946–) and Paul Perron (1940–) introduced the expression ‘conceptual metonymy’ as a parallel to conceptual metaphor as a way to organize reality in different cultures (1999: 176). Conceptual metonyms are another meaning-making tool at our disposal. Among the categories they suggest are the following ones.

PRODUCER FOR THE PRODUCT

1. Give me a *Perrier*®.
2. I just bought a *Ferrari*®.
3. I own a *Miró*.

THE PLACE FOR THE INSTITUTION

1. The *White House* just denied complicity.
2. *Fleet Street* is predicting a recession.
3. *Vatican City* issued a papal decree.

During the decade of the 1980s, Lakoff and Johnson presented a persuasive model of metaphor to explain how we convert concrete experience into abstract concepts and store the information gleaned from

those experiences in our brains. In the next section, we shall discuss how Lakoff and Johnson described this process.

The Image-Schema The image schema is a key element in the theoretical construct of Lakoff and Johnson's conceptual metaphor. Johnson (1987: 29) pointed out that an image schema “...operates at a level of mental organization that falls between abstract and propositional structures ... and particular concrete images. ...” Johnson argued that in order for us to provide structure to our experience and to be able to recall it, there has to be a pattern and order in our perceptions, actions, and conceptualizations. This arrangement occurs through image schemas that consist of shapes, patterns, and images that provide regularity to our experience. Schemas allow us to categorize and catalogue our experienced reality by associating the unknown with the known and allowing us to decide where the new information belongs. The experimental psychologist, Eleanor Rosch (1973), showed that people do this on a regular basis. In this way, conceptual metaphor, and its related forms conceptual metonymy and conceptual synecdoche, represent this cognitive process.

Johnson pointed out that the image schema is not a photographic reproduction of experienced reality; rather, it is a mental icon that is malleable in its format, i.e., is not a fixed form. It allows us to fit new experiences into a somewhat amorphous pattern or shape and give it meaning. Image schemas may represent any of the five senses (sight, hearing, touch, taste, smell). Selected linguistic analogues include the following expressions:

1. A sight to behold (sight)
2. The roar of the crowd (sound)
3. A clammy feeling (touch)
4. A sweet flavor (taste)
5. The smell of greasepaint (smell).

Because image schemas are such an ingrained part of language, we often do not recognize them.

Lakoff's (1987: 122–124, 138–142) work on fuzzy grammar pointed out systematic grammatical reflexes of what would come to be called image schemas. Lakoff showed that English utilizes certain phrases and expressions to express membership in prototypical categories. Thus, there are certain objects that clearly belong to a specific category, e.g., birds (winged, feathers, ability to fly) such as cardinals and bluejays. Other animals either do or do not belong to this category or belong only marginally. In these cases, expressions such as *a kind of*, *sort of*, *essentially*, and many more are employed to

situate these objects at the periphery of this category. Thus, the sentence *a penguin is a kind of a bird* indicates our placement of this animal at the border of this image-schematic classification. Ronald Langacker would develop this notion to a far greater extent in his ‘cognitive grammar’ model of linguistic theory (see below).

Lakoff and Johnson’s innovative research thus represents a continuum in the analysis of metaphor dating back to Aristotle. They utilize pre-existing scholarship as the basis for their conceptual metaphor theory. At the beginning of the 20th century, researchers such as Karl Bühler (1879–1963) initiated experimental research in the paraphrasing of metaphor-rich proverbial language. By the middle of the 20th century, psychologists including Charles Osgood (1916–1991), B. F. Skinner (1904–1990) and Solomon Asch (1907–1996) began to examine metaphor. It was I. A. Richards and Max Black, however, who most influenced Lakoff and Johnson’s research with their interactionist view of metaphor.

Ronald W. Langacker (b. 1942)

The theoretical linguist, Ronald W. Langacker, articulates Lakoff and Johnson’s views on conceptual metaphor in his cognitive grammar. Langacker notes that the Lakoff and Johnson view of metaphor is one in which metaphor is basic to epistemology. As stated previously, metaphor involves mapping a source domain (vehicle = the known) and a target domain (topic or tenor = the unknown). In this approach, the projection of a source field onto a target field is a metaphoric process. In this sense, as Danesi and Perron (1999: 175) noted that metaphor is “...the ability of the human brain to convert experience into abstraction via the mapping of some source domain onto a target domain to produce an abstract concept.” In his extensive research, Langacker provides specific examples from various languages to illustrate how his version of cognitive grammar works. What Langacker does is connect the conceptual metaphors to grammar. Specifically, he relates grammatical categorization and concept formation. The following example illustrates Langacker’s line of argumentation.

Lakoff demonstrates that certain aspects of English grammar reflect the image schema, the essential basis for Lakoff and Johnson’s notion of the conceptual metaphor. Langacker (2002: 1–32) showed a relationship between grammar and concept formation. In cognitive grammar, a metaphorical theory of grammar, we receive input from the external world. This input is processed at the experiential level, i.e., through our senses. Next, this information is

converted into image schemas such as shapes and so forth. Third, the image schemas are transformed into conceptual metaphors/metonyms/synecdoches. Finally, this totality of information is transformed into the linguistic categories of grammaticalized or lexicalized forms.

A succinct, albeit simplified restatement of one example, of the many presented by Langacker (2002: 149–163) suffices here. The verb *go* exemplifies the grammaticization, or the projection of certain basic concepts into the grammar of a language. The verb *go* has the meaning of motion and, subsequently, it evolves into a future meaning. This occurs because the original meaning of ‘movement from the speaker,’ a spatial concept, is transferred to a temporal domain. On the one hand, this verb means to move from point A to point B, e.g., $A \rightarrow B$. On the other hand, it signals a more abstract movement through time, i.e., an activity that will be carried out at a future time. In this sense, English and many other languages of the world combine spatial and temporal movement into a single verb. To be sure, this explanation is a bare-bones discussion of the complexities involved this semantic shift from space to time, but it serves to show that conceptual frameworks may be grammaticized or shifted from experience to an abstract grammatical notion of movement.

Literal Versus Figurative Language

The assertion that figurative language derives from a basic literal language has been a matter of discussion for some time. This underlying assumption thus separates language into two distinct categories – one primary, the other secondary. For nearly 25 centuries, since Aristotle’s commentary, the assumption of the literal-figurative language dichotomy has gone virtually unchallenged, albeit certain philosophers such as Vico have indeed insinuated that all language is metaphorical. As noted above, Aristotle’s significant statements about metaphor held sway, with certain precursory and insinuating statements to the contrary, until the 20th century.

The essence of the literal-figurative debate revolves around whether or not metaphor is a deviation from some pristine ordinary language or whether it is a basic form of linguistic expression. The essential question relates to whether or not metaphor is derivative or basic. To respond to this question, it is necessary to consider briefly some hypotheses about the origin of language. In this regard, Danesi (1993: 1–29) discussed in some detail the likely, lengthy progression of the process of linguistic evolution. Danesi’s response was Vichian in that he subscribed to Vico’s empiricist view that language originates

through gestures that signified basic concepts. Subsequently, these isolated gesticulations and primitive interjections became codified into an ever-increasing abstract set of oral symbols that we call language. This process involves an abstraction or a metaphorization of these primitive experiences that evolved from gesture to symbolic utterance to a fully developed linguistic syntax. Many of these primal metaphors become unrecognizable and they now appear to be literal linguistic expressions. Because these basic metaphors have become so commonplace and are now unrecognizable, we assume that they are literal.

In the literalist camp, Mac Cormac (1985: 73) defined literal language as “the use of language to express concrete objects and events. When we employ ordinary words in their ordinary dictionary senses to describe events or situations that are publicly perceptible, we are speaking literally.” Even though Mac Cormac is an ardent literalist who believes that there is a distinction between these two forms of language, he conceded that as a literal language apologist, he (Mac Cormac, 1985: 78) must admit that his explication of literal language presupposed an elementary metaphor, namely, that the world consists of natural categories.

In his 1936 discussion of metaphor, Richards was among the first to break with tradition and state that metaphor is essentially omnipresent in language. By making this bold statement, he revolutionized prevailing wisdom about metaphor. It is no longer deviation from standard or literal language, it is, in fact, the essence of language. Richards noted that the boundary between literal and figurative language was not fixed or constant.

Empirical research carried out by Pollio *et al.* (1977), for example, provided persuasive evidence for what Richards called the omnipresent nature of metaphor in ordinary language. Specifically, their research showed that individual speakers uttered approximately 3000 new metaphors per week, as well as 7000 idioms that are frozen metaphors.

Lakoff disputed the literal-figurative dichotomy. In an important article on this aspect of language, Lakoff (1986: 292) pointed out various distinct meanings of literal language. It is possible to state Lakoff's views of the misconceptions of the word ‘literal’ in this debate in at least four different ways. First, literal language diverges from special uses of language such as literary language. Second, there is a language for a particular subject matter. Third, there is a language that has no secondary meaning. Finally, there is a logical language that may be true or false. Lakoff points out that these four basic assumptions of a literal-figurative dichotomy are not tenable. In fact, the four meanings of ‘literal’ converge.

Semiosis and the Signifying Order

The signifying order is the means by which a society or a culture codifies meaning. It is a tripartite process, as Danesi and Perron (1999: 70) pointed out. It consists of semiosis (the capacity to comprehend and create signs), representation (the process of using signs to refer to objects and concepts), and the signifying order (the cultural complex employed by a given culture to impose representation and order in that society).

Sebeok (1920–2001) developed a modeling system to describe this process of the comprehension and production of signs (Sebeok, 1994: 116–127). The primary modeling system is the mental system that receives input from the external world. This system is the one that underpins the representational operations of replication, simulation, imitation, and indication. Next, the secondary modeling system is that mental system that produces projections of primary or concrete models into abstractions. In this sense, language exemplifies this modeling system because it utilizes primary signs to create linguistic signs. The tertiary modeling system extends primary and secondary modeling systems into an elaborate schema of cultural signs (literature, the performing and representational arts) to its most abstract and significant levels. The combination of these three systems, and most especially the tertiary modeling system is metaphorical, in nature because the cultural signs stand for different aspects of the culture, e.g., an Italian opera is a higher order system of signs (verbal, musical, artistic) that incorporates many essential aspects of the Italian culture.

Metaphor production is a special example of this process that involves, as Danesi and Perron (1999: 69) discussed, a series of stages that includes environmental input into the bodily senses through a process of semiosis. This input is then represented in the mind and ultimately displayed in the culture through the signifying order. In language and in art, this signifying order manifests itself through metaphoric concepts. The actual activity of sign-making is re-presentation, i.e., the presentation knowledge and information again in some new symbolic format, e.g., writing, painting, sculpture, architecture, and so forth. There is an ordered process for semiosis, namely, external input (through the senses) → re-presentation (the use of signs to refer to these external stimuli) → culture (the culture in which person lives gives meaning to the signs of that culture). One of the roles of educating a child is to teach it the salient signs of a culture. Parents engage in this activity with their children from infancy on. The signs of a specific culture are both verbal (language) and nonverbal (kinesics,

paralanguage, proxemics). Cultural signs are transmitted from one generation to another via literature, and the plastic and representational arts. Oral societies use memorized and recollected narratives, a kind of auditory cultural history to transmit their cultural signs from one generation to another. These codified signs must be learned by members of other cultures. In the deaf community often via various forms of translation, these signs take the form of meaningful systematic visual gestures.

The essence of semiotics, as Sebeok (1994: 11) pointed out, is the relationship expressed in the medieval Latin expression *aliquid stat pro quo*, i.e., 'something stands for something else.'

Charles Sanders Peirce would expand this bipartite definition (object-sign) of the sign to a tripartite one, in which interpretation becomes the third element (object-sign-interpretant). In essence, a sign is generally considered to consist of three components: (1) the sign or representation (something employed to stand for something else); (2) the object or referent (what is referred to by the sign); and (3) the interpretant (a person's understanding of the sign-object relationship, determined by culture). Medical analysis, one of the oldest recorded professional practices, provides a good example of this triadic system of meaning. In classical medicine, the four signs of illness were *fervor* 'fever,' *dolor* 'pain,' *rubor* 'redness,' and *tumor* 'swelling.' These signs or symptoms, alone or in combination, stand for various physical maladies to be interpreted by the physician. Over the years, symptomatology has become quite sophisticated, in part, aided by new knowledge and sophisticated technology. A good example of the medical sign is the aggregation of symptoms that stand for a stroke, namely, sudden numbness or weakness of the face or limbs, especially on one side, confusion, vision problems, and so forth. All of the medical signs converge to form a systematic pattern of a particular medical problem. The correct interpretation of this cluster of meaningful signs results in a diagnosis. The grouping of physical signs stands for specific illness or syndromes as interpreted by a trained physician.

The Uses of Metaphor

Metaphors may generate new knowledge. As noted above, this is often the case in science. The ability to explain the unknown in terms of the known, the abstract in terms of the concrete, make them ideal vehicles to expand knowledge and to enhance comprehension.

In any discussion of metaphor as a knowledge creator, especially in science, it is necessary to discuss three basic notions: (1) analogy, (2) model, and

(3) metaphor. It is not uncommon to find these terms used interchangeably, though, in fact, they are distinct. Analogy, a term used by Aristotle in the definition of his fourth type of metaphor, has at least two meanings. First, it may mean similarities in two objects that are otherwise distinct. Secondly, it is a kind of inference by which it may be deduced that objects that bear some similarity may also be comparable in other ways. A model in the scientific sense does not refer to a scale model, but rather to an analogy. A metaphor in science is again a comparison of the unknown in terms of the known. Thus, a miniature planetary system stands for the Bohr-Rutherford atom, a container of billiard balls in motion for the kinetic movement of gases, the computer for the brain, and so forth.

A model may be an object, often of reduced scale, that stands for the actual object. In architecture, scale models of buildings are commonplace. A second meaning is a tentative theoretical construct intended to function as a testing device. In her discussion of 'model,' Mary Hesse (b. 1924) uses the example of the dynamic model of gases, in which gases are seen as an assortment of billiard balls in random motion. In this model, motion and impact are the shared features, while color, hardness and shininess are irrelevant (Hesse, 1966: 7–11). The shared elements are a positive analogy; those that are not significant constitute a negative analogy. Other factors that play no role in the analogy are neutral. In this view of a 'model,' the primary model (model₁) alludes to the imperfect copy (billiard balls) minus the negative analogy. The second use of the term 'model' (model₂) is the second copy with its positive, negative, and neutral analogies.

A few examples of a model are in order. The Bohr-Rutherford conceptualization of the atom is explained in terms of the planetary system, i.e., the different parts of the atom with its component electrons and protons which function much as the planets that revolve around a central solar system. In this case, the planetary system is model₁ and the Bohr-Rutherford atom is model₂. Additional examples of this use of modeling include comparisons of the human or animal body to a machine, or the brain to a computer.

We shall now discuss a number of specific examples of how science and education use metaphors and models as a way to advance knowledge. It must be noted that a model is an actual object (planetary system, billiard balls, computers) already known to the audience. The theory used to explicate the scientific innovation uses the model metaphorically, once again, to describe the unknown in terms of the known.

Theory-Constitutive Metaphors

Scientists use metaphors to explicate their theories in a way that will make them comprehensible to their intended audience. This public, however, is often limited to fellow scientists. Successful metaphors, i.e., those that introduce a novel perspective in a branch of science, tend to capture the imagination of an entire generation of scholars who embrace its basic meaning and apply its newly created insights to previously unresolved problems. This notion of theory-constitutive metaphor merits review at this juncture.

Thomas S. Kuhn's (1922–1996) book *The structure of scientific revolutions* (1970) argued that there are periods of normal science when scientists work within a single paradigm or model commonly accepted by the practitioners of a particular science until there is what many call a 'breakthrough.' In his discussion of paradigm, Percival (1976: 286–287) noted that scientific paradigms consist of symbolic generalizations, such as Einstein's $e = mc^2$, which are elegant, yet simple explanations for previously unexplained phenomena. Paradigms also consist of models or conceptual analogies that offer members of a given scientific community an ontological framework. Members of these groups share theoretical values to evaluate competing theories within their framework. Finally, they use exemplars to show how their theory resolves complex problems that earlier theories could not. Periodically, this normal science is rejected, most often by a single scientific genius who provides answers to residual problems in the prior theory as well as making significant new strides in the field of inquiry. What we shall see is that scientific revolutions are often founded on novel metaphors that reject those of the theory they have supplanted.

Linguistics 20th-century linguistic theory provides an excellent example of the use of metaphor to change theory. During the last century, two major theoretical paradigms in linguistics emerged. The first was exemplified by Leonard Bloomfield's *Language* (1933), with its empiricist view of language. In this book, linguistic knowledge derives from external sensorial experience. In *Verbal behavior* (1957), B. F. Skinner, the U.S. psychologist, elaborated these empiricist views on language through his stimulus-response approach, one that greatly influenced second-language education in the 1960s. A second theoretical approach to linguistic analysis emerged with Noam Chomsky's *Syntactic structures* (1957) and *Aspects of theory of syntax* (1965), and its clearest exposition in Chomsky's *Cartesian linguistics* (1966), with its rationalistic approach to language

acquisition, i.e., humans are born with an inherent language acquisition device.

Both of these linguistic theories employ distinct metaphors. In brief, the Bloomfield model, or American Structural Linguistics, language may be viewed as a stack of boxes that represent levels of analysis (phonology, morphology, syntax). The linguist begins with the lowest box, or level, and then, presumably utilizes a series of analytical tools (item and arrangement analysis, item and substitution analysis) until the highest level (syntax) is attained. It utilizes a down-up metaphor. This approach utilizes the empiricist notion that language is accessible and analyzable through the senses. Structuralism likewise introduces language as a physical, biological, organic entity. Hence, phonemes function as atoms, syllables as molecules, and words as cells, though it was rare that structural analysis went beyond these linguistic units to higher order linguistic units such as sentences and discourse. Again, analysis starts with the phoneme, or basic unit of sound in a language, since its metaphoric model were those sciences which progressed from the smallest unit to the largest one. The conceptual metaphors that guided structural linguistic analysis, as we shall see, are very different from that those organized transformational-generative linguistic theory.

One of the problems with the structuralist model, and one that the next generation of linguists would address in a comprehensive fashion, is its failure to provide an analysis of syntax. In the succeeding generation, the transformational-generative model of language, Noam Chomsky viewed language as an abstract system genetically embedded in the brain. In this model, syntax is central, while phonology and morphology are lower-level, somewhat superficial manifestations of the abstract syntactic level. In this theoretical model, the brain functions as a computer, and it generates an infinite number of grammatical sentences via a limited set of simple rules. This model employs a left-right metaphor for analysis (rewrite rules). Transformational grammar also uses the metaphor of a tree structure with terminology such as branching, pruning, chopping, and nesting to describe its processes. Finally, geometrical and mathematical metaphors predominate, e.g., mapping, transforming, and propositional calculus. It is this use of mathematical metaphors that caused some to believe that the transformational-generative model of linguistic theory was a branch of mathematics. In the transformational model, ideas are often described as 'parallel,' 'diametrically opposed,' and 'central.' The fact that Western culture has a predisposition to accept mathematical and scientific models of the world as objective and authentic is another reason

for the switch from structuralism to a generative-transformational model of language.

An essential element of Chomsky's theory of language involves the computer metaphor. Several scholars observed that the computer metaphor is one that prevails when speaking of the mind or brain. Boyd (1979) examined this usage in his analysis of cognitive psychology. Boyd (1979: 360–361) made several equations including the following: (1) thought = information processing; (2) brain language = computations; (3) memory = data storage capacity; and (4) consciousness = feedback. Many of these same analogies exist implicitly in Chomsky's early work such as *Syntactic structures* (1957) and explicitly in his later work *The minimalist program* (1995) where, in fact, his language comes directly from computational jargon. It must also be noted that there are significant differences between the human brain and a computer as well as limitations on both.

Linguistics utilizes theory-constitutive metaphors to elucidate its basic principles. When there are significant changes in these metaphors, there is a paradigm change in the way linguistic theory examines its basic topic of study – language.

Physics Physics uses metaphor to explain and illustrate theories that would otherwise be difficult to comprehend. In his discussion of the metaphoric notion *boojum* (Mermin, 1990), N. David Mermin (b. 1935) points out that this expression derives from Lewis Carroll's (pseudonym of Charles Lutwidge Dodgson, 1832–1898) nonsense poem 'The hunting of the snark.' What Mermin wanted to do was describe the physical properties of a certain type of liquid at very low temperatures. At these low temperatures, the expected patterns could not be maintained, and as was the case with the snark, the physical behavior "softly and suddenly vanished away" (Carroll, 1939: 778). The neologism and the concept are now an established part of physics, as evidenced by its appearance as an entry in *Webster's new international dictionary*. Mermin's use of this metaphor captures the essence of the physical properties of this phenomenon.

Anatomy Sir William Harvey (1578–1657), whose anatomical research revealed the circulation of blood throughout the body, employed a mechanistic metaphor to illustrate his findings. In his description of the auricles and ventricles of the heart, Harvey likened them to the firing of a projectile from a firearm. To be sure, his description is incomplete but compelling.

Second Language Pedagogy Second language pedagogy is replete with metaphoric constructs, many of

them implicit. For Herron (1982), second language methodology consisted of a series of shifting metaphoric models. The grammar-translation model, for example, employs the MIND-BODY metaphor, by which second language acquisition entailed a sort of mental exercise akin to bodybuilding. A second metaphorical model, THE PRODUCTION METAPHOR, was characteristic of audiolingualism. In this approach, the process of imparting a second language to a student is similar to a factory in which a supervisor (instructor) molds the products (students) through their blueprints (textbooks). The basic purpose of this approach was a utilitarian one, in which use of the fundamental language skills (speaking, writing, comprehension, and reading) were the primary goals. The production metaphoric model still lingers in career-oriented classes, i.e., language for specific purposes (medicine, law, social work, and so forth). Herron also noted that a new metaphoric model has emerged; namely, that second language acquisition is the same as first language acquisition. One effect of this metaphor is to make the teacher a parent in the classroom, and this frequently results in a teacher-dominated classroom environment. Metaphoric models, especially unrecognized ones, may have a negative effect on second language instruction because they bear assumptions that may be counterintuitive to the results sought. Danahy (1986) concurred in this assertion and categorized metaphoric models for second language instruction and student-teacher interaction using major divisions – human-non-human metaphors and human-human ones. In the former case, the culinary metaphor entails that an instructor is the actor and the students are the patients. In this active/passive model, students are the recipients of a prepared meal that will enrich and enhance them. In the latter model, the family metaphor is one that assumes that students are child-like individuals incapable of doing anything on their own. This parent-child metaphor is a powerful that prevents students from becoming independent and self-sufficient. Recognition of the metaphoric models that pervade a discipline is essential to an understanding of what we do. Identification of negative or inappropriate conceptual metaphors in any profession is a first step to remedying deficiencies.

Concluding Remarks

In this overview of figurative language, we have seen that there are two basic theories of metaphor: the comparison theory that dates back to Aristotle, and its subsidiary version, the substitution model. In the 20th century, the interactionist theory, originally articulated by I. A. Richards and Max Black, but perhaps most completely developed and elaborated by

George Lakoff and Mark Johnson, provides us with a cognitive-conceptual model of metaphor and its corollaries metonymy and synecdoche to form a novel epistemological model.

See also: Aristotle and Linguistics; Empiricism; Language as an Object of Study; Metaphor: Philosophical Theories; Modern Linguistics: 1800 to the Present Day; Philosophy of Linguistics.

Bibliography

- Aristotle (1997). *Aristotle's poetics*. Whalley G (trans.). Montreal & Kingston: McGill-Queen's University Press.
- Black M (1981 [1955]). 'Metaphor.' In Johnson M (ed.) *Philosophical perspectives on metaphor*. Minneapolis, MN: University of Minnesota. 63–82.
- Boyd R (1979). 'Metaphor and theory change: what is 'metaphor' a metaphor for?' In Ortony A (ed.) *Metaphor and thought*. Cambridge: Cambridge University Press. 356–408.
- Carroll L (1939). *The complete works of Lewis Carroll*. New York: Modern Library.
- Danahy M (1986). 'On the metaphorical language of L2 research.' *The Modern Language Journal* 70, 228–235.
- Danesi M (1993). *Vico, metaphor, and the origin of language*. Bloomington, IN: Indiana University Press.
- Danesi M & Perron P (1999). *Analyzing cultures: an introduction and handbook*. Bloomington, IN: Indiana University Press. (This is an excellent introduction to cultural semiotics that provides information on topics discussed herein.)
- Eco U (1984). *Semiotics and the philosophy of language*. Bloomington, IN: Indiana University Press.
- Groupe μ (1981). *A general rhetoric*. Burrell P B & Slotkin E M (trans.). Baltimore: Johns Hopkins University Press.
- Herron C (1982). 'Foreign language learning approaches as metaphor.' *The Modern Language Journal* 66, 235–242.
- Hesse M B (1966). *Models and analogies in science*. Notre Dame, IN: Notre Dame University Press.
- Johnson M (1987). *The body in the mind: the bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.
- Kuhn T S (1970). *The structure of scientific revolutions* (2nd edn.). Chicago: University of Chicago Press.
- Lakoff G (1986). 'The meanings of literal.' *Metaphor and Symbolic Activity* 1, 291–296.
- Lakoff G (1987). *Fire, women and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff G & Johnson M (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Langacker R W (2002). *Concept, image, and symbol: the cognitive basis of grammar* (2nd edn.). Berlin: Mouton de Gruyter.
- Mac Cormac E R (1985). *A cognitive theory of metaphor*. Cambridge, MA: MIT Press.
- Mermin N D (1990). *Boojums all the way through: communicating science in a prosaic age*. Cambridge: Cambridge University Press.
- Noppen J-P & Hols N (eds.) (1990). *Metaphor 2: a classified bibliography of publications from 1985–1990*. Amsterdam: John Benjamins.
- Noppen J-P, Knop S de & Jongen R (eds.) (1985). *Metaphor: a bibliography of post-1970 publications*. Amsterdam: John Benjamins.
- Percival W K (1976). 'The applicability of Kuhn's paradigms to the history of linguistics.' *Language* 52, 285–299.
- Pollio H, Barlow J, Fine H & Pollio M (1977). *The poetics of growth: figurative language in psychology, psychotherapy, and education*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Richards I A (1936). *The philosophy of rhetoric*. Oxford: Oxford University Press.
- Ricoeur P (1977). *The rule of metaphor: multi-disciplinary studies of the creation of meaning in language*. Czerny R (trans.). Toronto: University of Toronto Press.
- Rosch E (1973). 'Natural categories.' *Cognitive Psychology* 7, 532–547.
- Schroots J J F (1991). 'Metaphors of aging and complexity.' In Kenyon G M, Birren J E & Schroots J J F (eds.) *Metaphors of aging in science and the humanities*. New York: Springer.
- Sebeok T A (1994). *Signs: an introduction to semiotics*. Toronto: University of Toronto Press.
- Shibles W A (1971). *Metaphor: an annotated bibliography and history*. Whitewater, WI: Language Press.
- Vico G (1968). *The new science of Giambattista Vico*. Bergin T G & Fisch M H (trans.). Ithaca, NY: Cornell University Press.
- Wellek R & Warren A (1956). *Theory of literature* (3rd edn.). New York: Harcourt Brace & World.

Formal Semantics

G Chierchia, Università degli Studi di Milano-Bicocca, Milan, Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Semantics, in its most general form, is the study of how a system of signs or symbols (i.e., a language of some sort) carries information about the world. One can think of a language as constituted by a lexicon (an inventory of morphemes or words) and a combinatorial apparatus according to which complex expressions, including, in particular, sentences, can be built up. Semantics deals with the procedures that enable users of a language to attach an interpretation to its arrays of symbols. Formal semantics studies such procedures through formally explicit mathematical means.

The history of semantics is nearly as long and complex as the history of human thought; witness, e.g., the early debates on the natural vs. conventional character of language among the pre-Socratic philosophers. The history of formal semantics is nearly as daunting as it is intertwined with the development of logic. In its modern incarnation, it is customary to locate its inception in the work of logicians such as Frege, Russell, and Tarski. A particularly important and relatively recent turning point is constituted by the encounter of this logico-philosophical tradition with structural and generative approaches to the study of human languages, especially (though by no means exclusively) those influenced by N. Chomsky. The merger of these two lines of research (one brewing within logic, the other within linguistics), has led formal semantics to become a central protagonist in the empirical study of natural language. The research paradigm that has emerged has proven to be quite fruitful, both in terms of breadth and depth of results and in terms of the role it is playing in the investigation of human cognition. The present work reviews some of the basic assumptions of modern formal semantics of natural language and illustrates its workings through a couple of examples, with no pretence of completeness.

Semantics vs. Lexicography

One of the traditional ideas about semantics is that it deals with the meaning of words. The main task of semantics is perceived as the compilation of dictionaries (semantics as lexicography). To this, people often add the task of investigating the history of words. Such a history can teach us about cultural development. One might even hope to arrive at the

true meaning of a word through its history. Compiling dictionaries or reconstructing how particular words have changed over time are worthy tasks; but they are not what formal semantics is about. Lexicography, philology, and related disciplines vs. semantics as conceived here constitute complementary enterprises. They all, of course, deal with language. But the main goal of semantics is to investigate how we can effortlessly understand a potential infinity of expressions (words, phrases, sentences). To do that, we have to go beyond the level of single words.

It may be of use to point to the kind of considerations that have led semantics to move the main focus of investigation away from single word meanings and their development. For one thing, it can be doubted that word histories shed light on how words are synchronically (i.e., at a given point in time) understood and used. People use words effectively in total ignorance of their history (a point forcefully made by one of the founding fathers of modern linguistics, namely F. de Saussure). To make this point more vividly, take the word *money*. An important word indeed; where does it come from? What does its history reveal about the true meaning of money? It comes from Latin *moneta*, the past participle feminine of the verb *moneo* 'to warn/to advise.' *Moneta* was one of the canonical attributes of the Roman goddess Juno; *Juno moneta* is 'the one who advises.' What has Juno to do with money? Is it perhaps that her capacity to advise extends to finances? No. It so happens that in ancient Rome, the mint was right next to the temple of Juno. So people metonymically transferred Juno's attribute to what was coming out of the mint. A fascinating historical fact that tells us something as to how word meanings may evolve; but it reveals no deep link between money and the capacity to advise. This example is not meant to downplay the interest of historical investigations on word meanings; it is just an illustration of how linguistic history affects only marginally the way in which a community actually understands its lexicon.

There is a second kind of consideration suggesting that the scope of semantics cannot be confined to the study of word meanings. Do words in isolation have clearly identifiable meanings? Take any simple word, say the concrete, singular, common noun *dog*. What does it mean? Some possible candidates are: the dog-kind, the concept of dog, the class of individual dogs. . . . And the list can go on. How do we choose among these possibilities? Note, moreover, that all these hypotheses attempt to analyze the meaning of the word *dog* by tacking onto it notions (kind, concept, class . . .) that are in and of themselves in need of

explication. If we left it at that, we wouldn't go far. Looking at dictionary definitions is no big help either. If we look up the entry for *dog*, typically we will find something like:

- (1) A highly variable carnivorous domesticated mammal (*Canis familiaris*) prob. descended from the common wolf.

Indeed, if someone doesn't know the meaning of the word *dog* and knows what *carnivorous* and *mammal* mean, then (1) may be of some practical help. But clearly to understand (1), we must rely on our understanding of whole phrases and the words occurring in them. Words which, in turn, need a definition to be understood. And so on, in a loop. This problem is sometimes called the problem of the circularity of the lexicon. To put it differently, (1) is of help only if the capacity to use and interpret language is already taken for granted. But it is precisely such capacity that we want to study.

The limitation of a purely word-based perspective on the investigation of meaning is now widely recognized. Frege summarized it in a nice motto: "only in the context of a sentence do words have meaning." His insight is that complete sentences are linguistic units that can sort of stand on their own (more so than any other linguistic units). They can, as it were, express self-contained thoughts. We are more likely, therefore, to arrive at the meaning of single words (and of phrases in between words and complete sentences) via a process of abstraction from the contribution that words make to sentence meaning, rather than the other way around. This is so because sentence meaning is somehow more readily accessible (being, as it were, more complete) than the meaning of words in isolation.

These are some reasons, then, why the perspective of modern semantics is so different from and complementary to lexicography and philology; such perspective is much more directly tied to the investigation of the universal laws of language (language universals) and of the psychological mechanisms underlying such laws. Understanding the function, use, etc., of a single word presupposes a whole, complex cognitive apparatus. It is, therefore, an arrival point more than a starting point. It seems thus reasonable to start by asking what it is to understand a sentence.

The main thesis we wish to put forth is that to understand a sentence involves understanding its relations to the other sentences of the language. Each sentence carries information. Such information will be related to that of other sentences while being unrelated to that of yet others. In communicating, we rely on our spontaneous (and unconscious) knowledge of these relations.

The Notion of Synonymy and Its Problems

Imagine watching a Batman movie in which the caped hero fights the Riddler, one of his eternal foes. The Riddler has scattered around five riddles with clues to his evil plans. Batman has managed to find and solve four of them. We could report this situation in any of the following ways:

- (2a) Batman has found all of the five clues but one.
 (2b) Batman has found four out of the five clues.
 (2c) Four of the five clues have been found by Batman.

These sentences are good paraphrases of each other. One might say that they have roughly the same information content; or that they describe the same state of affairs; or that they are (nearly) synonymous. (I will be using these modes of speaking interchangeably.) To put it differently, English speakers know that there is a tight connection between what the sentences in (2a), (2b), and (2c) mean. This is a kind of knowledge they have *a priori*, i.e., regardless of what actually goes on. Just by looking at (2a) vs., say, (2b) and grasping what they convey, we immediately see that they have roughly the same informational content.

This is what we mean when we say that understanding a sentence involves understanding which other sentences count as good paraphrases and which don't. Thus, knowing a language is to know which sentences in that language count as synonymous. Semantics is (among other things) the study of synonymy. Two synonymous sentences (and, more generally, two synonymous expressions) can always be used interchangeably. This last informal characterization can be turned into a precise definition along the following lines.

- (3a) Suppose one utters any complex expression α containing a subexpression A. If one can replace in A α with a different expression β , without changing the overall communicative import of A, then α and β are synonymous.
 (3b) α is synonymous with β = in the utterance of any expression A containing α , α can be replaced with β without changing the communicative import of the utterance (*salva significatione*).

For example, in uttering (2a) (our A), we can replace the subcomponent that comes after *Batman has found*—namely *all of the five clues but one* (our α) with *four out of the five clues* (our β) and convey exactly the same information. Hence, these two expressions must be synonymous (and, in fact, so are the whole sentences).

This looks promising. It paves the way for the following setup for semantics. Speakers have intuitions of whether two expressions can be replaced

with each other while keeping information content unchanged. For any two sentences α and β , they spontaneously know whether they can be substituted for each other (i.e., whether β can be used to paraphrase α). Because the sentences of a language are potentially infinite, it is impossible for speakers to memorize synonymous sentences one by one (for that clearly exceeds what our memory can do). Hence, they must recognize synonymy by rule, by following an algorithm of some sort. The task of semantics, then, becomes characterizing such an algorithm.

There is a problem, however. Sameness of communicative import is a more or less thing, much like translation. In many contexts, even sentences as close as those in (3a) and (3b) could not be replaced felicitously with each other. Here is a simple example. The discourse in (4a) is natural and coherent. The one in (4b) much less so:

- (4a) Batman has found all of the five clues but one,
which is pinned on his back.
(4b) ?? Batman has found four out of the five clues,
which is pinned on his back.
(modeled after a famous example by B. Partee)

Clearly in (4a) we cannot replace *Batman has found all of the five clues but one* with *Batman has found four out of the five clues* while keeping unaltered the overall communicative effect. This means that if we define synonymy as in (3a) and (3b), then (2a) and (2b) cannot be regarded as synonymous after all. Yet they clearly share a significant part of their informational content. What *is* it that they share?

In fact, it has been argued that if (3a) and (3b) are how we define synonymy, then there simply are no two sentences that qualify as such. Here is a classical argument that purports to show this (based on Mates (1950)). Take the following two sentences:

- (5a) Billy has a dog.
(5b) Billy has a highly variable carnivorous
domesticated mammal prob. descended from
the common wolf.

Are these two sentences synonymous? Hardly. They are clearly semantically related. But they surely do not have the same communicative import. Nor can one replace the other in every context. For example, (5a) could describe a true state of affairs, while (5b) might not:

- (6a) Molly believes that Billy has a dog.
(6b) Molly believes that Billy has a highly variable
carnivorous domesticated mammal prob.
descended from the common wolf.

This shows that in contexts like *Molly believes that* — we cannot simply replace a word with its dictionary

definition. And if dictionary definitions don't license synonymy, then what does?

The problem can be couched in the following terms. Any normal speaker of English perceives a strong semantic connection among the sentences in (2a), (2b), and (2c), or (4a) and (4b). So strong that one might feel tempted to talk about synonymy. Yet when we try to make the notion of synonymy precise, we run into serious problems. Such a notion appears to be elusive and graded (a more or less thing); so much so that people have been skeptical about the possibility of investigating synonymy through precise, formal means.

A fundamental breakthrough has been identifying relatively precise criteria for assessing semantic relations. The point is that perfect synonymy simply does not exist. No two sentences can be always replaced with each other. The notion of synonymy has to be deconstructed into a series of more basic semantic relations. We need to find a reliable source for classifying such relations, and, we will argue, such a source lies in the notions of truth and reference. Consider the sentences in (2a), (2b), and (2c) again. Assume that the noun phrase *the five clues* in (2a) and (2b) refer to the *same* clues (i.e., we are talking about a particular episode in a particular story). Then, could it possibly happen that say (2a) is true and (2b) false? Evidently not: no one in his right mind could assert (2a) while simultaneously contending that (2b) is false. If (2a) is true, (2b) also must be true. And, in fact, vice versa: if (2b) is true, then (2a) also must be. When this happens, i.e., when two sentences are true in the same set of circumstances, we say that they have the same truth conditions.

Notice that sameness of truth conditions does not coincide with or somehow require sameness of communicative import (too elusive a notion), nor substitutivity in any context whatsoever (a condition too difficult to attain). Our proposal is to replace such exceedingly demanding notions with a series of truth-based notions, while keeping the same general setup we sketched in connection with synonymy: for any pair of sentences, speakers have intuition about whether they are true under the same conditions or not. They can judge whether they are true in the same (real or hypothetical) circumstances or not. Because the sentences of our language are infinite, this capacity must be somehow based on a computational resource. Speakers must be able to compare the truth-conditions associated with sentences via an algorithm of some sort. The task of semantics is to characterize such an algorithm. The basic notion changes (synonymy is replaced with sameness of truth conditions), but the setup of the problem stays the same.

Truth and Semantic Competence

Let us elaborate on the proposal sketched at the end of the previous section. Information is transmitted from one agent to another (the ‘illocutionary agents’) in concrete communicative situations (‘speech acts’). No two such situations are alike. And consequently, no two pieces of information that are transmitted through them are alike. In *Groundhog Day*, a movie with the actor Bill Murray, the protagonist gets trapped into going through the same day over and over. He wakes up and his day starts out in the same way (with the alarm clock ringing at 7 a.m. on groundhog day); as he walks outside, he meets the same waitress who greets him in the same way (“weather so-so, today”). Yet this sentence, though being the same day after day, and being uttered in circumstances as identical as they can conceivably be, clearly conveys a different sense or information unit on each occasion of its use (the hearer going from noticing that something is fishy about this verbatim repetition, to the painful discovery of the condemnation to live through groundhog day for eternity).

Ultimately, we want to understand how communication takes place. But we cannot nail down every aspect of a speech act, just as we cannot know (not even in principle, I believe) every aspect of the physical or mental life of a particular human being. At the same time, while speech acts are unique events, there is much that is regular and invariant about them; that is what can be fruitfully investigated. One family of such invariants concerns form: similar sound patterns may be used in different speech acts. Another family of invariants concerns content: similar states of affairs may be described through a variety of expressions. The notion of truth is useful in describing the latter phenomenon. A pair of sentences may be judged as being necessarily true in the same circumstances. This is so, for example, for (5a) vs. (5b). Yet, such sentences clearly differ in many other respects. One is much more long-winded than the other; it uses rarer words, which are typical of high, formal registers. So in spite of having the same truth conditions, such sentences may well be used in different ways. Having the same truth condition is generally regarded as a semantic fact; being able to be used in different ways is often regarded as a pragmatic fact. While this gives us a clue as to the role of these two disciplines (both of which deal with meaning broadly construed), the exact division of labor between semantics and pragmatics remains the object of controversy.

Truth conditions are a tool for describing semantic invariants, structural regularities across communicative situations. Whenever I utter a declarative sentence, I typically do so with the intention to

communicate that its truth conditions are satisfied (which of course raises the question of nondeclaratives, emotive expressions, and the like; see for example textbooks such as Chierchia and McConnell-Ginet (2000) or Heim and Kratzer (1998); cf. also Kratzer (1999) for a discussion of relevant issues). Truth conditions depend on the reference (or denotation) of words and the way they are put together, i.e., they are compositionally projected via the reference of words (or morphemes). If I say to you, as we are watching a movie, “Batman has found all of the five clues but one,” you understand me because you sort of know (or guess) who Batman is, what sort of things clues are, what finding something is, what number the word *five* refers to; you also understand the “all ... but ...” construction. The reference/denotation of words is set (and is modified, as words may change their denotation in time) through use, in complex ways we cannot get into within the limits of the present work. The denotation of complex expressions (e.g., of a verb phrase such as [_{VP}found five clues] and truth conditions of sentences) are set by rule (the semantic component of grammar). Semantic rules presumably work like syntactic rules: they display variation as well as a common core, constitutive of universal grammar. Insofar as semantics is concerned, what is important for our purposes is that truth conditions can be compositionally specified. This paves the way for an algorithmic approach to meaning. We already remarked that sentences are formed by composing morphemes together via a limited number of syntactic operations. So to arrive at the truth condition of an arbitrary sentence, we can start by the contribution of the words (their reference). Then, for each way of putting words together, there will be a way of forming the reference of complex expressions, and so on until we arrive at the truth condition of the target sentence.

So far, we have discussed sentences that have the same truth conditions (such as those in (2a), (2b), and (2c)); but this is not the only semantic relation that can be characterized in terms of the notion of truth. Consider the following examples.

- | | |
|--------------------------|-------------------------|
| (7a) Every Italian voted | a'. Most Italians voted |
| for B | for B. |
| (7b) Leo voted for B | b'. Leo voted for B. |

Sentence (7a) is related to (7b) in a way that differs from the relation between (7a') vs. (7b'). Here is the difference. If (7a) is true, and Leo is Italian, then (7b) has to be true, too; this is clearly not so for (7a') vs. (7b'): (7a') may be true without (7b') being true. If whenever A is true, B also must be, we say that A *entails* B (B's meaning is part of A's meaning). Two sentences with the same truth conditions entail

each other (i.e., they hold a symmetric relation); when entailment goes only one way (as from (6a) to (6b)), we have an asymmetric relation.

Entailment is pervasive. Virtually all semantic intuitions are related to it. As an illustration, consider the pair of sentences in (8a) and (8b).

- (8a) John promised Bill to take him to the station.
- (8b) John ordered Bill to take him to the station.

Pronouns, like *him* in (8a) and (8b), take their denotation from the context; they can take it from the extra linguistic context (a person salient in the visual environment, a person the speaker points at, etc.) or from the linguistic context (e.g., from NPs that occur in the same discourse; *John* or *Bill* in (8a) and (8b)); one widespread terminology is to speak of *indexical* uses in the first case and of *anaphoric* uses in the second. We can conceptualize this state of affairs by viewing pronouns as context-dependent items, incomplete without pointers of some sort. Now we shall focus on the anaphoric interpretation of (8a) vs. (8b).

- (9a) John promised Bill that John would take Bill to the station.
- (9b) John ordered Bill that Bill should take John to the station.

These appear to be the only options. That is to say, sentence (8a) cannot convey something like 'John promised Bill that Bill should take John to the station.' The point of this example is that we have intuitions that govern how the denotation of a pronoun is to be reconstructed out of contextual clues; such intuitions tell us that (8a) and (8b), though structurally so similar, allow for a distinct range of interpretive options. At the basis of intuitions of this sort, we again see entailment at work: on its anaphoric construal, (8a) entails (9a).

Another important set of truth-based semantic relations are presuppositions. Consider the contrast between the sentences in (10a) and (10b).

- (10a) Fred stole the cookies.
- (10b) It was Fred who stole the cookies.

There is a noticeable semantic contrast between (10a) and (10b). How can we characterize it? Clearly the two sentences are true in the same circumstances (they entail each other). Yet they differ semantically. Such a difference can be perhaps caught by looking at what happens by embedding (10a) and (10b) in a negative context.

- (11) So, what happened this morning?
- (11a) Everything went well. Fred didn't steal the cookies; he played with his toys.
- (11b) ?? Everything went well. It wasn't Fred who stole the cookies.

The answer in (11a) is natural. The one in sentence (11b) would sound more natural as an answer to

- (12a) Who stole the cookies?
- (12b) It wasn't Fred.

The difference between the question in (11) and the one in (12a) is that the latter (but not the former) tends to presuppose that cookies were stolen. In other terms, the situation seems to be the following. Both sentences in (10a) and (10b) entail:

- (13) Someone stole the cookies.

If either (11a) or (11b) are true, then (13) must also be true. Furthermore, sentence (13) must be true for (10b) to be *denied* felicitously. The illocutionary agents must take for granted the truth of (13) to assert, deny, or otherwise use sentence (10b), as the naturalness of the following continuations for (13) illustrate:

- (14) Someone stole the cookies ...
- (14a) It was Fred.
- (14b) It wasn't Fred.
- (14c) Was it Fred?
- (14d) If it was Fred, he is going to get it ...

This brings us to the identification of presupposing as a distinctive semantic relation: a sentence A presupposes B if the truth of B must be taken for granted in order to felicitously assert, deny, etc., A. Presuppositions are quite important in language. So much so that there are distinctive syntactic constructions (such as those in (10b), known as cleft sentences) specifically keyed to them.

Let me illustrate the wealth of semantic relations and their systematic character by means of another example, which will bring us to the interface between semantics and pragmatics. Consider:

- (15a) Who stole the cookies?
- (15b) Fred looks mischievous.
- (15c) Fred stole the cookies.

If to a question such as (15a), I reply with (15b), I do suggest/convey something like (15c). Sentence (15c) clearly is not part of the literal meaning of (15b) (however hard defining such a notion might be). Yet, in the context of the dialogue in (15a), (15b), and (15c), speakers will converge in seeing that (15c) is strongly suggested by (15b). Here, too, we have, thus, a systematic semantic intuition. The suggestion in (15c) can be retracted; that is, one can continue (15b) with '... but I know he didn't do it'. However, in the absence of such an explicit correction, illocutionary agents upon hearing (15b) will tend to infer (15c). This phenomenon has been studied by H. P. Grice (1989), who dubbed it implicature. His

proposal is that it arises through interaction of the core meaning assigned to sentences by rule with principles that govern conversational exchanges. The basic idea is that for conversational exchanges to be successful they have to be basically cooperative acts; cooperating means that one sticks to relevant topics, one only gives information believed to be truthful, one gives no more and no less than what is relevant, etc. Applying this to the case at hand, in a situation in which question (15a) is topical, answering (15b) would seem to be blatantly irrelevant; the hearer, however, tends to interpret it as relevant and sets in motion an inferential process that tends to link it to some piece of information that does address the topical question; such a link is to be found with the help of the information available in the context to the illocutionary agents (e.g., in the common knowledge that if people commit a mischief, such as stealing cookies, they may well look mischievous, etc.). Thus, this type of semantic judgment (the implicature) appears to be best accounted for in terms of the interaction between grammar and general conditions on reasonable language use (that fall under the scope of pragmatics).

Sometimes it is not immediately clear whether something is a matter of conventionalized meaning or pragmatics. To illustrate, consider the oscillation in meaning of a word like *or*. It can be illustrated with the following examples. Consider first (16a):

- (16a) If I got it right, either John or Mary will be hired.
- (16b) If I got it right, either John or Mary but not both will be hired.

Normally, one tends to interpret (16a) as truth conditionally equivalent to (16b); i.e., the disjunction in (16a) is interpreted exclusively (as incompatible with the simultaneous truth of each disjunct). However, this is not always so. Contrast (16a) with (17a).

- (17a) If either John or Mary are hired, we'll celebrate.
- (17b) (?) If John or Mary (but not both) are hired, we'll celebrate.
- (17c) If John or Mary or possibly both are hired, we'll celebrate.

The most natural interpretation of (17a) is not the exclusive one (namely (17b), which is somewhat odd pragmatically); rather it is the inclusive one, made explicit in (17c). (Notice that the emphatic word *either* is present both in (16a) and (17a); in spite of this, the interpretation of *or* shifts.) We might see in these phenomena a lexical ambiguity of disjunction. Words expressing disjunction, we may feel inclined to conclude, have a varying interpretation, as it happens

with words such as *bank* or *lap* ('sit on my lap' vs. 'he swam three laps'). We may assume that such interpretations are always in principle available, but then we select the most suitable to the context of the speech act. While this seems *prima facie* possible, there are reasons to doubt it. In particular, true lexical ambiguities are resolved across languages (in Italian, there are two different words for the two senses of *lap*). Ambiguities are never universal. The meaning shift of *or*, *per contra*, seems to be universal: in every language disjunction appears to have a similar oscillation in meaning. A convincing case for two lexically distinct disjunctions, one exclusive, the other exclusive, has not been made (sometimes it has been proposed that Latin *vel* vs. *autem* is just that; for arguments against this, cf., e.g., Jennings (1994)). Moreover, other areas of the lexicon have been found that display a similar behavior (e.g., the number words). This strongly suggests that a different explanation for such behavior should be found. Grice himself has proposed that the phenomenon under discussion is to be accounted for in terms of the interaction between semantics and pragmatics. The idea is that the basic meaning of *or* is the inclusive one, as it is the most liberal interpretation; the exclusive construal arises as an implicature, i.e., a pragmatic enrichment, albeit a generalized one. The advantage of this move is that it would explain the oscillation in meaning of disjunction without positing a covert ambiguity. We will come back to how the generalized implicature associated with *or* might come about in the later section "The Semantics/Pragmatics Interface".

Wrapping up, the picture that emerges is roughly the following. In using language, speakers display complex forms of spontaneous knowledge. They put together words in certain ways and not others. This is how knowledge of syntax manifests itself. They also accept certain paraphrases and not others, draw certain inferences and not others, etc. It turns out to be possible/useful to categorize the latter in three major families of semantic relations.

- (18a) Entailment-based (entailment, mutual entailment, contradictoriness, analyticity, etc.)
- (18b) Presupposition-based (presupposition, question/answer pairs, etc.)
- (18c) Implicature-based (generalized implicature, particularized implicature, etc.)

All of them can be readily defined in terms of the notion of truth:

- (19a) A entails B = for any conceivable situation *s*, if A is true in *s*, B is also true in *s*.

- (19b) A presupposes B = to use A appropriately in a situation *s*, the truth of B must be taken for granted by the illocutionary agents in *s*.
 (19c) A implicates B = use of A in a situation *s* suggests, everything else being equal, that B is true in *s*.

The definitions in (19a), (19b), and (19c) can be readily associated with “operational” tests that enable speakers to assess whether a given relation obtains or not. For example, to check whether (20a) entails (20b), you might check whether you could sincerely assert (20a) while denying (20b), viz. whether you could sincerely and felicitously utter something like (20c):

- (20a) It is indeed odd that Mary is home.
 (20b) Mary is home.
 (20c) It is indeed odd that Mary is home, even if she in fact isn't.

To the extent that you can't really say something like (20c), you are entitled to conclude that (20a) entails (20b). It is useful, in these cases, to use contrast sets such as (21a) and (21b).

- (21a) It is indeed conceivable that Mary is at home.
 (21b) It is indeed conceivable that Mary is home, even if she in fact isn't.

The semantic relations in (18a), (18b), and (18c) can be viewed as intuitions of semantic relatedness speakers have about sentences of their own language, as judgments that may be elicited, and the like. By analogy with well-formedness judgments, there are some cases in which things are not so clear and we may not be sure whether, say, a certain entailment holds or not. In such a case, more complex arguments, indirect evidence of various sorts, or psycholinguistic experimentation may be called for (see, e.g., Crain and Thornton (1998) on experimental methodologies for truth-based semantic judgments). But in indefinitely many cases, simple introspection yields relatively straightforward judgments. The capacity for making such judgments is constitutive of our semantic competence. Such a competence cannot be simply a thesaurus, a store of pairs of sentences, with the relative judgment tacked on, for the number of judgments speakers can make on the fly is potentially infinite. Semantic competence must be a computational device of some sort. Such a device given an arbitrary pair of sentences $\langle A, B \rangle$ must be able to determine in principle whether A entails B, presupposes it, etc. The task of semantics is to characterize the general architecture of such a computational device. While there are many foundational controversies that permeate the field, there is a broad convergence that this is roughly the

form that the problem of meaning takes within modern formal semantics.

Semantic Modeling

In the present section I will sketch how a (necessarily, much simplified) calculus of semantic relations may look. Suppose you have a lexicon of the following form:

- (22a) N: John, Bill, dog, cat, table, . . .
 (22b) V: runs, smokes, drinks, . . .
 (22c) DET: the, a, some, every, no . . .

Think of syntax as a device that combines lexical entries by merging them in complex phrases and assigning them a syntactic analysis that can be represented by tree diagrams or labeled bracketings of the following form:

- (23a) [_{VP} John smokes]
 (23b) [_{DP} every boy]
 (23c) [_{VP} [_{DP} every boy] smokes]

I assume, without being able to justify it, that lexical items have phrasal projections. In particular, VP is the phrasal projection of V and constitutes a clausal nucleus composed of the verb and its arguments linked in a predicative structure. Such a nucleus forms the innermost skeleton of the sentence (I will have to ignore matters pertaining to inflection, agreement, tense, and the like). The lexical features of verbs are crucial in determining the characteristics of clausal nuclei. DP is the phrasal projection of D, and it is constituted by a determiner and a (common) noun. Clausal nuclei can be formed by merging a verb with a (proper) name or a DP, as indicated. In the spirit of the discussion in the section on **Truth and Semantic Competence**, semantics assigns recursive truth conditions to sentences in terms of the reference assigned to lexical entries. There are several ways to do this. Ultimately, the choice one makes on the exact format of interpretive rules has far-reaching consequences for our understanding of grammar. However, our choices here are only in small part dictated by our current understanding of semantics in universal grammar; for the major part, they result from considerations such as ease of exposition, keeping prerequisites at a minimum, and the like. To get started, we should assign a reference (or denotation, terms we will use interchangeably) to lexical entries. To do so, we assume we have a certain domain $D_s = \{a, b, c, \dots\}$ at each given discourse situation *s* that constitutes our universe of discourse. A discourse situation can be thought of as the time at which the utterance takes place. A domain is just a set of individuals,

pragmatically selected (e.g., those salient to the illocutionary agents). Interpretations are relative to an utterance situation s and the corresponding domain of discourse D_s . Reference of proper nouns, for example, is suitably chosen from the domain of discourse. Suppose, for example, that a and b are salient humans in our universe of discourse, then we might have:

- (24) For any conceivably relevant utterance situation s , the name *John* denotes a in s ; the name *Bill* denotes b in s ...

It doesn't matter how a or b are characterized (via a description, an act of indication, etc.) to the extent that one successfully succeeds in linking the noun to its bearer. Also, it is useful to have a uniform category-neutral notation for semantic values; we will use for this the double bar notation $\| \cdot \|$; accordingly, for any expression α , $\|\alpha\|^s$ will be the semantic value of α in situation s . Thus, (24) can be abbreviated as:

- (25) $\|John\|^s = a$ (where $a \in D_s$, the domain of discourse at s)

(Technically, $\| \cdot \|$ can be viewed as a function from expressions and situations into denotations; so sometimes we will speak of the interpretation function.) The denotation of a simple (intransitive) verb such as those in (22b) can be thought of as a function that for each (appropriate) individual in the domain discourse tells us whether that individual performs a certain action or not. Here is an example:

- (26) *smokes* in a situation s denotes a function smoke_s that applies to animate individuals and returns truth values. If a is such an individual, then $\text{smoke}_s(a)$ returns 'true' (which we represent as the number 1) if that individual performs the action of smoking in s (where smoking involves ...); otherwise $\text{smoke}_s(a)$ returns 0 (i.e., 'false').

If a is not animate (e.g., if a is a stone and s is a 'normal' situation), then $\text{smoke}_s(a)$ is not defined (lacks a value). The final part in definition (26) reflects the fact that sentences like (27a) and (27b), out of the blue, are (equally) strange: smoking normally requires its subject argument to be animate.

- (27a) That stone smokes.
(27b) That stone doesn't smoke.

The deviance of sentences like (27a) and (27b) has been variously characterized as a violation of selectional restrictions or as sortal deviance. Here we are couching the relevant phenomenon in presuppositional terms (to illustrate a further application of such a concept). The fact that sentences of this sort remain deviant across negation may be taken as

evidence that the verb *smoke* imposes an animacy presupposition on its arguments (see e.g., Chierchia and McConnell-Ginet (2000) for more discussion). A definition like (26) can be stated more compactly:

- (28) $\|\text{smokes}\|^s = \text{smoke}_s$,
where for each a in D_s , $\text{smoke}_s(a)$ is defined iff a is animate in s ; if defined, $\text{smoke}_s(a) = 1$ if a smokes in s (where smoking involves ...);
 $\text{smoke}_s(a) = 0$, otherwise.

The definition of (or constraints on) smoking (i.e., the dots in (28)) can be elaborated further in several ways by refining our lexical analysis of the verb *smoke*. Although much progress has been made on this score, many important issues remain open (including, e.g., whether a presuppositional treatment of selectional restrictions is ultimately viable). What is important, from the point of view of compositional semantics, is the *logical type* or *semantic category* of the denotation of a verb like *smoke*. Such verbs are treated here as functions from individuals into truth values. These are called *characteristic functions*; they divide the (relevant portion of) the domain of discourse of the utterance situation in two: the things that satisfy the verb from those that don't. Characteristic functions correspond to sets (which might be called the *extension* of the function), as the following example illustrates:

- (29) Let universe of discourse be $\{a, b, c, d\}$; let a, b , and c be people. Of these, let a and b smoke in s while b but not a also smokes in a different situation s' . We can represent all this as follows:
- | | |
|---------------------------------------|-------------------------------------|
| $a \rightarrow 1$ | |
| $\text{smoke}_s = b \rightarrow 1$ | corresponding extension: $\{a, b\}$ |
| $c \rightarrow 0$ | |
| $a \rightarrow 1$ | |
| $\text{smoke}_{s'} = b \rightarrow 0$ | corresponding extension: $\{a\}$ |
| $c \rightarrow 0$ | |

As is evident from the example, sets and characteristic functions are structurally isomorphic (encode the same information). In what follows it will be useful on occasion to switch back and forth between these two concepts. Use of characteristic functions as a formal rendering of verb meanings is useful in giving truth conditions for simple subject predicate sentences:

- (30a) A sentence of the form $[_{VP} N V]$ is true in s iff $\|V\|^s (\|N\|^s) = 1$
Example:
(30b) $[_{VP} \text{Bill drinks}]$ is true in s iff $\|\text{drinks}\|^s (\|\text{Bill}\|^s) = 1$

The truth conditions of any sentence with the syntactic structure specified in (30a) boil down to applying a

characteristic function to an individual (and thereby ascertaining whether that individual belongs to the set that constitutes the extension). To find out whether Bill in fact smokes in *s*, we need factual information about the situation obtaining in *s*. To understand the sentence, we don't. We merely need to know its truth conditions, which in the case of simple subject-predicate sentences are an instruction to check the value of a characteristic function for the argument specified by the subject. The rules in (30a) and (30b) can be reformulated more compactly as in (31):

$$(31) \llbracket [_{VP} N V] \rrbracket^s = \llbracket V \rrbracket^s (\llbracket N \rrbracket^s)$$

This can be viewed as the kernel of a predication rule (that tells us how subject and predicates combine semantically).

Everything so far looks like a formally explicit (and perhaps somewhat pedantic) way of sketching a denotational, information-oriented semantics, and the reader may get the feeling of not yet finding striking insights on what meaning is. In order to grasp the potential of this method, one needs to look at a little more of its computational apparatus. So let us turn now to DPs. Things here are definitely more challenging. DPs are constituents formed by a determiner plus a common noun. Common nouns can be given, at least in first approximation, the same analysis as (intransitive) verbs, i.e., the meaning of, say, *cat* can be thought of as a characteristic function that selects those entities that are cats out of the universe of discourse (or, equivalently, we can say that *cat* identifies a class/set across situations). But what about things like *no cat* or *every cat*, which are the typical constituents one finds in, e.g., subject position and the like? What does *no cat* denote? And, even worse, what do *no* or *every* or *some* denote? Our program is to assign a denotation to lexical entries and then to define in terms of it truth conditions for sentences. So we must find suitable denotations of Ds and DPs.

To address questions of this sort, we apply a heuristic that goes naturally with our general setup: whenever the denotation of an expression is not directly accessible to your intuition, look at what that expression contributes to the truth conditions of the sentences it occurs in (the epistemological primacy of sentences, again). So, consider for example:

(32) No boy smokes.

We know/assume/conjecture that *boy* and *smoke* denote characteristic functions and that sentences contribute truth values (i.e., they are true or false, as the case may be, in different situations). We may think of *no* as a function, too. As is evident from

(32), such a function combines first with a characteristic function/set (corresponding to the noun); then the result combines with a second characteristic function (corresponding to the verb) to yield a truth value. Schematically, here is what we have:

$$(33) \text{no}(\text{boy}_s) (\text{smoke}_s) = 1 \text{ or } 0$$

Now we can look at our intuitions. When is (32) true? The answer is pretty clear. When among the boys, nobody smokes. Or, equivalently, when the class of boys (i.e., the extension of *boy_s*) has no member in common with the smokers (i.e., the extension of *smoke_s*), (32) is true. In set talk, the intersection between the boys and the smokers must be empty:

$$(34) \text{no}(\text{boy}_s) (\text{smoke}_s) = 1 \text{ iff } \text{BOY}_s \cap \text{SMOKE}_s = \emptyset$$

(where *BOY_s*, *SMOKE_s* are the extensions corresponding to *boy_s*, *smoke_s*, respectively) This is perfectly general. Replace boy/smokes with any other noun/verb. The contribution of *no* stays constant: *no*(*N*) (*V*) is true just in case no member of the extension of *N* is in *V*. We thus discover that *no* has a perfectly sensible (if abstract) denotation: a function that encodes a relation between sets. Our contention here is that speakers behave as if they had such a function in mind (or something similar to it) in using *no*.

The next step is to see that *all* determiners express relations among sets (characteristic functions), just like *no* does. Here are a few examples, along with some comments.

(35a) *Some*

(35a.i) Example: some boy smokes

(35a.ii) Truth conditions: *some*(*boy_s*) (*smoke_s*) = 1 iff *BOY_s* ∩ *SMOKE_s* ≠ ∅

(35a.iii) Comment: *some* is the contrary of *no*; *some boy smokes* is true just in case you can find someone among the boys who is also among the smokers; i.e., the intersection between the class of boys and the class of smokers must be non empty. The indefinite article *a* can be analyzed along similar lines.

(35b) *Every*

(35b.i) Example: every boy smokes

(35b.ii) Truth conditions: *every*(*boy_s*) (*smoke_s*) = 1 iff *BOY_s* ⊆ *SMOKE_s*

(35b.iii) Comment: *every* expresses the subset relation: *every boy smokes* is true just in case all the members of the class of boys also belongs to the class of smokers

(35c) *Most*

(35c.i) Example: Most boys smoke

(35c.ii) *most*(*boy_s*) (*smoke_s*) = 1 iff the number of member of *BOY_s* ∩ *SMOKE_s* is

bigger than half the number of members of BOY_s .

- (35c.iii) Comment: *most* involves actual counting. *Most boys smoke* is true just in case the number of boys who smoke (i.e., the intersection of the boys with the smokers) is greater than half the number of boys (i.e., more than half of the boys are smokers).

(35d) *The*

- (35d.i) Example: The blond boy smokes.

- (35d.ii) Truth conditions: $\text{the}(\text{blond boy}_s)(\text{smoke}_s)$ is defined only if there is exactly one blond boy in s . Whenever defined, $\text{the}(\text{boy}_s)(\text{smoke}_s) = \text{every}(\text{boy}_s)(\text{smoke}_s)$.

- (35d.iii) Comment: this reflects the fact that *the blond boy smokes* is only interpretable in situations in which the universe of discourse contains just one blond boy. If there is more than one blond boy or if there is no blond boy, we wouldn't really know what to make of the sentence. So *the* is a presuppositional determiner; it presupposes the existence and uniqueness of the common noun extension. (This analysis of *the* goes back to Frege.)

In spite of the sketchiness of these remarks (that neglect important details of particular determiners), it should be evident that the present line of analysis is potentially quite effective. A class of words and phrases important and tendentially stable across many languages falls into place: determiners ultimately express natural relations between sets (the set associated with the common noun and the set associated with the verb phrase). Our denotational perspective seems to meet rather well the challenge that seemingly denotationless items pose. It is useful to see what becomes of our rule of predication (viz. (31) above). Evidently such a rule needs to be split into two (main) subcases, depending on whether the subject is a simple N (a proper name) or a complex DP. Here is an exemplification of the two cases:

(36a) Mary smokes.

(36b) No boy smokes.

In case (36a), we have semantically two pieces: an individual (whomever *Mary* denotes) and a characteristic function (smoke_s); so the latter applies to the former. In case (36b) the two pieces are: a complex function (namely $\text{no}(\text{boy}_s)$) that looks for a characteristic function to yield a truth value, and, as before, the characteristic function smoke_s ; in this case the former applies to the latter. In either case, the end result is a truth value. So our predication rule becomes:

(37a) $\| [\text{VP N V}] \|^\S = \|\text{V}\|^\S (\|\text{N}\|^\S)$

(37b) $\| [\text{VP DP V}] \|^\S = \|\text{DP}\|^\S (\|\text{V}\|^\S)$

This suggests that the core rule of semantic composition is functional application. Consider for example an ungrammatical sentence of the form:

(38) * $[\text{VP boy smokes}]$

Such a sentence, as things stand, would be generated by our (rudimentary) syntax. However, when we try to interpret it, we find two characteristic functions of individuals, neither of which can apply to the other. Hence, the sentence is uninterpretable, which explains its ungrammaticality. There are languages like, for example, Russian or Hindi where singular common nouns without a determiner can occur in subject position:

(39a) Russian: mal'cik kurit boy smokes 'the boy smokes'

(39b) Hindi: kamre meN cuuha ghuum rahaa hai (from Dayal 2004) room in mouse moving is 'a mouse is moving in the room'

Notice that (39a) is the verbatim translation of (38) and is grammatical in Russian. The line we are taking suggests that in such languages it must be possible to turn some covert forms of common nouns into argumental DPs, i.e., things that can semantically combine with predicates; for example it is conceivable that in a language without articles, like Russian, the semantic functions associated with the articles can be applied covertly (as part of the interpretive procedure), so as to rescue the semantic mismatch that would otherwise ensue. This may, in turn, involve the presence of a phonologically null determiner (for alternative developments of this line of analysis, as well as details concerning the available interpretations, see, e.g., Chierchia (1998), Longobardi (2001), and Dayal (2004)).

The picture that emerges is the following. The basic mode of syntactic composition is *merge*, or some analogously simple operation that puts together two constituents (subject to parametrization pertaining to, e.g., word order, case, etc.). The basic mode of semantic composition is *apply*: constituents are compositionally analyzed as functions (of more or less complex semantic type) and arguments (individuals or other functions); so whenever we find a function and an argument of the appropriate sort, we simply apply the former to the latter. If things go wrong at any level, the derivation crashes and the result is ungrammatical. The semantic side of this process has come to be known as 'type driven interpretation,' the main idea being that the semantic categories of functions and arguments drive the interpretation process.

The present approach directly yields a computationally tractable theory of entailment and presupposition.

We have defined entailment roughly as follows: a sentence *S* entails a sentence *S'* iff whenever *S* is true, *S'* is also true. The apparatus we have developed allows us to prove whether a certain entailment holds or not. Let me show, as an illustration, that (40a) entails (40b) but not vice versa.

- (40a) Every scientist smokes.
- (40b) Every mathematician smokes.

To show this we need to assume that if one is a mathematician, one is a scientist; i.e.,

- (41) For every individual *a*,
- (41a) if **mathematician**_s (*a*) = 1, then **scientist**_s (*a*) = 1 or, equivalently:
- (41b) **MATHEMATICIAN**_s ⊆ **SCIENTIST**_s

Consider now the semantics of (40a), according to our analysis. It is the following:

- (42) **every**(**scientist**_s) (**smoke**_s)

In virtue of (35b), this is tantamount to

- (43) **SCIENTIST**_s ⊆ **SMOKE**_s

This being so, every subset of the set of scientists must also be included among the smokers (by elementary set theoretic considerations). Since, in particular, mathematicians are scientists, it follows that

- (44) **MATHEMATICIAN**_s ⊆ **SMOKE**_s

But this is just the semantics of (40b). So, if (40a) is true in *s*, then (40b) must also be true in *s*. Evidently, this reasoning goes through no matter which situation we are in. Hence, (40a) does entail (40b). On the other hand, it is easy to conceive of a situation in which (44), and hence (40b), hold, but say some economist doesn't smoke; in such a situation, (43) would fail to obtain. Hence, (40b) does not entail (40a).

A fully parallel way of reasoning can be put forth for presuppositions. We said that *S* presupposes *S'* iff *S'* must be taken for granted in every situation in which *S* is asserted, denied, etc. This can be cashed in as follows. We can say that for *S* to be true or false (i.e., to have a semantic value that makes it suitable for assertion or denial), *S'* must be known to be true in the utterance situations by the illocutionary agents, i.e., *S* can be true or false in *s* iff *S'* is true in *s*. Using this definition (known as the 'semantic' definition of presupposition), we can formally prove (though we will not do so here) that, for example, (45a) presupposes (45b):

- (45a) The blond boy smokes.
- (45b) There is exactly one blond boy around.

The general point of these examples is the following. Intuitions about entailment and the like are

a priori; speakers have them just by inspecting the meaning of the relevant sentences. In the present setup, this central fact is captured as follows. Semantics can be viewed as a set of axioms that (a) determines the interpretation of lexical entries and (b) assigns truth conditions to sentences. Such apparatus yields a calculus of entailment (and other semantic relations) that reemerge as theorems of semantics. We have not formalized each single step of the derivation (relying on the readers' patience and understanding of elementary set theory); but such a formalization is, evidently, feasible. We not only thereby gain in clarity. We also obtain a device that constitutes a reasonable (and falsifiable) model of speakers' linguistic abilities. The claim is not that the specific rules we have given are actually implemented in the speakers' mind. The claim is that speakers, to the extent that they can be said to compute entailments must be endowed with computational facilities that bear a structural resemblance to the ones sketched here. This, in turn, paves the way for inspecting the architecture of our linguistic abilities ever more closely. Without excessive optimism and in full awareness of the controversies that permeate the field, this seems to constitute a step in the right direction.

One further remark on the general picture that emerges from the sketch above cannot be avoided. Our approach to meaning is denotational: we assign a denotation to words and morphemes and (in terms of such denotations) truth conditions to sentences. This can be understood in several ways, of which I will present two much simplified extremes. We can take truth condition assignment as a way of exposing the link between language and the world, which is, arguably, the ultimate goal of semantics. Words/morphemes are actually mapped into aspects of the world (e.g., names are mapped into actual individuals); sentences are symbolic structures that code through their fine structure how things may be arranged in the world. However, it is also possible to view things somewhat differently. What really matters, it can be argued, is not the actual mapping between words and aspects of reality and between sentences and the conditions under which they, in fact, are true. What we do is give a form or recipe or potential for actual truth conditions; we merely constrain the form that truth conditions may take. What we get out of this is what really matters: a calculus of semantic relations (entailment, presupposition, etc.). Unlike what happens in, say, pure logic, such a calculus is not a normative characterization of sound reasoning; it is an empirically falsifiable characterization of semantic competence (i.e., of what speakers take to follow from what, when). Under

the latter view, truth conditions (or truth condition potentials, or whatever it is that we map sentences on) are a ladder we climb on to understand the working of semantic relations, i.e., relations that concern the information content of linguistic expressions.

It is evident that we are not going to settle these issues here. As a small consolation (but also, if you wish, as evidence of the maturity of the field), I hope to have given the reader reasons to believe that progress is possible even if such foundational issues remain open.

We haven't discussed implicatures and other pragmatically driven intuitions about meaning. To understand the full scope of the present proposal, it is important to do so. This requires extending a bit what we have done so far.

The Semantics/Pragmatics Interface

In the section **Truth and Semantic Competence**, we mentioned implicatures, a broad and varied type of meaning relations. We will elaborate by looking more closely at the oscillation in the meaning of *or*. The purpose is to illustrate how generalized implicatures come about and how this bears on the view of semantics sketched in the preceding section on **Semantic Modeling**.

The first step is to attempt a semantic analysis of *or*. To this we now turn. Imagine we extend our grammar by introducing coordination and negation along the following lines:

- (46a.i) $[_{VP} \text{ John doesn't smoke}]$
- (46a.ii) $[_{VP} \text{ NEG VP}]$
- (46b.i) $[[_{VP} \text{ John smokes}] \text{ and/or } [_{VP} \text{ Bill smokes}]]$
- (46b.ii) $[_{VP} \text{ and/or VP}]$

The syntax of negation and coordination poses many thorny questions we simply cannot address here. Although for our purposes any number of assumptions concerning syntax might do, let us maintain, again without much justification, that a negative sentence like (46a.i) has the structure in (46a.ii) out of which the observed word order is derived by moving the subject left from the inner VP. Furthermore, we will assume that coordinated sentences, whether disjunctive or conjunctive, such as (46b.i), are obtained through schemas such as (46b.ii). Insofar as semantics is concerned, the introduction of negation, conjunction, disjunction, etc., poses problems similar to that of determiners. The relevant expressions are function words, and it is not obvious how to analyze them in denotational terms. This question, however, can be addressed in much the same way as we have done with the determiners: by looking at what the relevant elements contribute to the truth conditions of the

sentences they occur in. For sentential operators, we can draw on a rich logical tradition. In the attempt to characterize the notion of valid inference, logicians have discussed extensively propositional connectives (like *not*, *and*, *or*), and the outcome is an analysis of such elements as truth functions or, equivalently, in terms of 'truth tables.' For example, the contribution of negation to meaning can be spelled out in terms of conditions of the following sort:

- (47a) *John doesn't smoke* is true in s iff *John smoke* is false in s

- (47b) $\| \text{NEG VP} \|_s = 1$ iff $\| \text{VP} \|_s = 0$

(47c)	VP	NEG VP
	1	0
	0	1

In (47c) we display in the form of a truth table the semantics given in (47b). Essentially, this says that in uttering a negation like (47a), the speaker intends to convey the falsity of the corresponding positive sentence. By the same token, conjunctions can be analyzed as in (48a), (48b), and (48c), and disjunction as in (49a), (49b), and (49c):

- (48a) *John smokes and Bill smokes* is true if both *John smokes* and *Bill smokes* are.

- (48b) $\| [VP_1 \text{ and } VP_2] \|_s = 1$ iff $\| VP_1 \|_s = \| VP_2 \|_s = 1$

(48c)	VP ₁	VP ₂	$[VP_1 \text{ and } VP_2]$
(48c.i)	1	1	1
(48c.ii)	1	0	0
(48c.iii)	0	1	0
(48c.iv)	0	0	0

- (49a) *John smokes or Bill smokes* is true if either *John smokes* or *Bill smokes* or both are true.

- (49b) $\| [VP_1 \text{ or } VP_2] \|_s = 1$ iff either $\| VP_1 \|_s = 1$ or $\| VP_2 \|_s = 1$ or both

(49c)	VP ₁	VP ₂	$[VP_1 \text{ or } VP_2]$
(49c.i)	1	1	1
(49c.ii)	1	0	1
(49c.iii)	0	1	1
(49c.iv)	0	0	0

This is the way in which such connectives are analyzed in classical (Boolean) logic. Such an analysis has proven extremely fruitful for many purposes. Moreover, there is little doubt that the analysis in question is ultimately rooted in the way in which negation, etc., works in natural language; such an analysis indeed captures at least certain natural uses of the relevant words. What is unclear and much debated is whether such an analysis stands a chance as a full-fledged (or nearly so) analysis of the semantics of the corresponding English words. There are plenty of cases where this seems *prima facie* unlikely. This is

so much so that many people have concluded that while Boolean operators may be distilled out of language via a process of abstraction, they actually reflect normative principles of good reasoning more than the actual semantics of the corresponding natural language constructions. Of the many ways in which this problem might be illustrated, I will choose the debate on the interpretation of *or*.

The interpretation of *or* provided in (49a), (49b), and (49c) is the inclusive one: in case both disjuncts turn out to be true, the disjunction as a whole is considered true. As we saw, this seems adequate for certain uses but not for others. The exclusive *or* can be analyzed along the following lines:

(50) Exclusive <i>or</i>			
	VP ₁	VP ₂	[VP ₁ or VP ₂]
(50.i) 1	1	0	
(50.ii) 1	0	1	
(50.iii) 0	1	1	
(50.iv) 0	0	0	

As the readers can verify by comparing (49a), (49b), and (49c) with (50), the two interpretations of *or* differ only in case (i); if both disjuncts are true, the whole disjunction is true on the inclusive interpretation and false on the exclusive one.

So, the thesis that *or* is ambiguous can be given a precise form. There are two homophonous *ors* in English. One is interpreted as in (48a), (48b), and (48c), the other as in (50). Illocutionary agents choose among these options on pragmatic grounds. They go for the interpretation that is best suited to the context. Determining which one that is will involve knowing things like the topic of the conversation (e.g., are we talking about a single job or more than one), the purpose of the conversational exchange, the intentions of the speaker, etc.

We mentioned that Grice proposed an alternative view, however. We are now in position to spell it out more clearly. If you look closely at the two truth tables in (49a), (49b), and (49c) vs. (50), you'll notice that in all the cases in which the exclusive *or* comes out true (namely case (ii) and case (iii)), the inclusive one does, too, i.e., in our terms, $[p \text{ or}_{\text{exclusive}} q]$ entails $[p \text{ or}_{\text{inclusive}} q]$. The former is, thus, stronger, more informative than the latter in the following precise sense: it rules out more cases. If you get the information that $[p \text{ or}_{\text{exclusive}} q]$ holds, you know that case (ii) or case (iii) may obtain, but case (i) and case (iv) are ruled out. If you know instead that $[p \text{ or}_{\text{inclusive}} q]$ obtains, you know that you might be in case (i), (ii), or (iii); only case (iv) is ruled out. Your degree of uncertainty is higher. So $\text{or}_{\text{exclusive}}$ is more restrictive; $\text{or}_{\text{inclusive}}$ is more general (more liberal we said). Things being so, suppose for a moment that *or* in

English is unambiguously inclusive (i.e., its interpretation is the most general, less restrictive of the two); this does not rule out at all the possibility that we are in case (ii) or case (iii). The exclusive construal, in other words, might arise as a special case of pragmatic strengthening. It is as if we silently add to, say, (51a) something like (51b).

(51a) John or Mary will be hired.

(51b) (... but not both)

The silent addition of (51b) to (51a) might be justified through a reasoning of the following sort:

(52) The speaker said (51a); let us assume she is being cooperative and not hiding on purpose any relevant information. This entails that she has no evidence that both John and Mary have been hired, for otherwise she would have said so. Assuming, moreover, that she is well-informed about the facts, this furthermore entails that she thinks that in fact (51b) holds.

So in this view, the base interpretation (viz. (51a)) is enriched through an inferential process that draws on principles of rational conversational exchanges and on factual knowledge about the context. The relation between (51a) and (51b) can thus be analyzed as a case of implicature (cf. on this, e.g., Horn (1989), Levinson (2000), and references therein).

The debate on how the two interpretations of *or* come about is important and shows different ways in which semantics is taken to interact with broader considerations pertaining to communication. Whether the two interpretations of *or* are a matter of ambiguity or arise as an implicature, I want to point out a generalization concerning their distribution, which I think shows something important concerning how language works. I will argue that the cases in which *or* is construed preferentially inclusively are (1) predictable, and (2) determined by structure. Then, I will put forth a hypothesis as to why this is so.

We have seen that a sentence like (16a), repeated here as (57a), is interpreted as in (57b), namely exclusively:

(53a) If I got it right, either John or Mary will be hired.

(53b) If I got it right, either John or Mary but not both will be hired.

Now take the consequent (i.e., the main clause) in the conditional in (53a) and move it to the antecedent, and the interpretation tends to shift:

(54a) If either John or Mary are hired, we'll celebrate.

(54b) If John or Mary or both are hired, we'll celebrate.

So, moving a disjunction from the consequent to the antecedent seems to have a systematic effect on the interpretation of *or*. The same holds for the pair in (55a) and (55b):

- (55a) Every student will either take an exam or write a paper.
 (55b) Every student who either takes an exam or writes a paper will satisfy the requirements.

In (55a), *or* is within the VP, which corresponds to the second argument of *every*, according to the analysis sketched in the section **Semantic Modeling**. Its preferred interpretation is exclusive. In (55b), *every* is in a relative clause which is part of the subject NP (namely, the first argument of *every* according to the analysis in **Semantic Modeling**). Its preferred interpretation is clearly inclusive.

A further class of contexts that displays a similar effect are negation and negative verbs. Compare (56a) and (56b):

- (56a) I believe that either John or Mary will be hired.
 (56b) I really doubt that either John or Mary will be hired.

Sentence (56a) is likely to get the interpretation ‘I believe that either John or Mary but not both will be hired.’ Sentence (56b), on the other hand does not have a parallel reading. It rather means ‘I really disbelieve that John and Mary stand a chance.’

The list could go on. But these examples should suffice to instill in the reader the idea that there is a systematic effect of structure on the interpretation of *or*. A doubt might linger, though, as to whether it is really in the nature of structure to have this impact. Take, for example, the pair in (55a) and (55b). Is it the position of disjunction that makes a difference? Or is it rather our knowledge of how classes normally work?

This is a legitimate question. Noveck *et al.* (2002) address it experimentally. They designed a reasoning task, in which logically naïve subjects are asked to judge whether a certain inference is sound or not. For example, subjects were asked to judge whether one can infer (57c) from (57a) and (57b):

- (57a) If there is an A, then there is a B or a C.
 (57b) There is an A.
 therefore:
 (57c) There aren’t both a B and a C.

Subjects were told that this was about inferences that could be drawn (on the basis of the given premises) concerning letters written on the back of a certain blackboard. What would your answer be? The experimental subjects overwhelmingly accepted the inference in (57a) and (57b). What is interesting is that in

terms of classical Boolean logic (which takes *or* to be inclusive) this inference is invalid. It is only valid if *or* in (57a) is interpreted exclusively. At the same time, subjects rejected inferences of the following form:

- (58a) If there is an A, then there is a B and a C.
 (58b) There is an A.
 therefore:
 (58c) There is a B or a C.

Again, this seems to make sense only if *or* in (58c) is interpreted exclusively. Things change dramatically if *or* is embedded in the antecedent of a conditional:

- (59a) If there is an A or a B, then there is a C.
 (59b) There is an A; there is also a B.
 therefore:
 (59c) There is a C.

Subjects overwhelmingly accepted this inference as valid. But this is only possible if *or* in (59a) is construed inclusively. Our raw intuition thus finds experimental confirmation, one that passes all due controls (the inferences were mixed with others containing other connectives and quantifiers, so that subjects were not conditioned to devise an answering strategy, and the order of presentation was duly varied, etc.). What is interesting is that these experiments only involved meaningless letters A, B, C... so scripts, contextual clues, knowledge of the world can hardly be imputed any role in the outcome. If there is a systematic effect on the interpretation of *or*, this must be due to the meaning of conditionals, of disjunction, and to the positioning of the latter. Nothing else is at play.

The reader may wonder how one manages to find out which structures affect the interpretation of *or*. The answer is that such structures were familiar from another phenomenon: the licensing of Negative Polarity Items (NPIs). NPIs are lexical items like *any* or *ever* that seem to require the presence of a negative element:

- (60a) * There is any cake left
 (60b) There isn’t any cake left.

NPIs are acceptable in the contexts that favor the inclusive interpretation of *or* over the exclusive one:

- (61a) * If we are in luck, there are any cookies left
 (61b) If there are any cookies left, we are in luck.
 (62a) * Everyone had any cookies left
 (62b) Everyone who had any cookies left shared them.

This correlation is striking, for the two phenomena (the distribution of *any* and of inclusive vs. exclusive *or*) seem to have little in common.

The next question is whether the relevant contexts have some common property. The answer seems to be

positive and, surprisingly, points in the direction of a rather abstract, entailment-based property. Positive contexts typically license inferences that go from sets to supersets. For example, (63a) entails (63b) and not vice versa.

- (63a) There are Marlboros.
- (63b) There are cigarettes.

The set of cigarettes is a superset of the set of Marlboros; so the entailment goes from a set to its supersets. Negation reverses this pattern: (64b) entails (64a) and not vice versa:

- (64a) There aren't any Marlboros.
- (64b) There aren't any cigarettes.

Now the VP portion of a sentence with *every* (i.e., its second argument) patterns with (63a) and (63b):

- (65a) Everyone had Marlboros.
- (65b) Everyone had cigarettes.

Sentence (65a) entails sentence (65b) and not vice versa. So does the consequent of a conditional

- (66a) If you open the drawer, you'll find Marlboros.
- (66b) If you open the drawer, you'll find cigarettes.

But the NP argument of *every* (its first argument) inverts this pattern just like negation, as we saw in the **Semantic Modeling** section:

- (67a) Everyone who had Marlboros shared them.
- (67b) Everyone who had cigarettes shared them.

Here it is (67b) that entails (67a) and not vice versa. The same applies to the antecedent of conditionals:

- (68a) If you smoke Marlboros, you'll be fined.
- (68b) If you smoke cigarettes, you'll be fined.

Sentence (68b) entails (68a); on the other hand (68a) could be true without (68b) necessarily being true (in a town in which Marlboros but no other brand is banned).

In conclusion, the contexts that favor the inclusive interpretation of *or* share a semantic property that has to do with entailment patterns: they all license entailments from sets to their subsets. Such a property has come to be seen as the property of being *downward entailing* (where *down* refers to the directionality of the entailment from sets to smaller ones). If this characterization is correct, this means that speakers to the extent that they interpret *or* as shown, must differentiate such contexts, and hence must be able to compute the entailments associated with the relevant structure.

The next question is why *or* tends to be interpreted inclusively in downward entailing structures. I will only hint at what strikes me as a highly plausible

answer. As we saw above, in plain unembedded contexts, exclusive *or* is stronger (i.e., asymmetrically entails) than inclusive *or*. The set of cases in which exclusive *or* is true is a subset of the set of cases in which the inclusive one is true. We evidently prefer, everything else being equal, to go for the strongest of two available interpretations. Now, negation and, in fact, all downward entailing structures, as we just saw, reverse this pattern. Under negation, first becomes last; i.e., strongest becomes weakest. In the case of disjunction, the negation of inclusive *or* is stronger (i.e., entails) the negation of exclusive *or*. I'll leave it to the readers to persuade themselves that this is so. Now why is this observation relevant? Suppose we go for the strongest of two alternatives (i.e., we maximize informativeness, everything else being equal); for disjunction, in downward-entailing contexts inclusive *or* is the strongest interpretation; in nondownward-entailing contexts exclusive *or* is the strongest. This explains the observed behavior in terms of a rather simple principle that optimizes information content on the basis of the available expressive resources.

So pragmatic strengthening (via a generalized implicature) correlates harmoniously with the entailment properties of various elements.

Conclusions

We have sketched a view of semantic competence as the implicit knowledge a speaker has of how the information content of various expressions is related. We have proposed to classify the hosts of semantic relations in three major families: entailment-based, presupposition-based, and implicature-based. Given two sentences, speakers can judge whether they entail each other or not, whether they presuppose each other or not, and so on; and they can do so with finite cognitive resources. We have sketched a denotational semantics that accounts for such a competence (i.e., provides a model for it). Our semantics takes the form of a calculus in which entailments, presuppositions, and even (certain) implicatures re-emerge as theorems. Such a model is formal in the sense of being explicit (building on the tradition of logic and model theory). It is, however, also substantive, in that it models a human cognitive capacity (i.e., the ability to semantically relate sentences to each other). We have seen two simple applications of this approach, to the analysis of determiners and connectives. We have also discussed a case of pragmatic enrichment. What we found is that the interpretation of *or* as exclusive or inclusive follows a pattern sensitive to downward entailingness (much like what happens with negative polarity items). If this is so, then

entailment patterns are not simply an invention of logicians or linguists. They must be constitutive, in an unconscious form, of the spontaneous knowledge that endows speakers with their linguistic abilities.

See also: Boole and Algebraic Semantics; Compositionality; Semantic Aspects; Implicature; Monotonicity and Generalized Quantifiers; Presupposition; Quantifiers: Semantics.

Bibliography

- Chierchia G (1998). 'Reference to kinds across languages.' *Natural Language Semantics* 6, 339–445.
- Chierchia G & McConnell-Ginet S (2000). *Meaning and grammar* (2nd edn.). Cambridge, Mass: MIT Press.
- Crain S & Thornton R (1998). *Investigations in Universal Grammar*. Cambridge, Mass: MIT Press.
- Dayal V (2004). Number marking and (in)definiteness in kind-terms. *Linguistics and Philosophy* 27, 393–450.
- Grice H P (1989). *Studies in the ways of words*. Cambridge, Mass: Harvard University Press.

- Heim I & Kratzer A (1998). *Semantics in generative grammar*. Oxford: Blackwell.
- Horn L (1989). *A natural history of negation*. Chicago: University of Chicago Press.
- Jennings R E (1994). *The genealogy of disjunction*. Oxford: Oxford University Press.
- Kratzer A (1999). Beyond ouch and oops. How expressive and descriptive meaning interact. Paper presented at the Cornell Conference on Context Dependency, unpublished manuscript. Amherst Mass.
- Levinson S (2000). *Presumptive meanings*. Cambridge, Mass: MIT Press.
- Longobardi G (2001). 'How comparative is semantics? A unified parametric theory of bare nouns and proper names.' *Natural Language Semantics* 9, 335–369.
- Mates B (1950). 'Synonymy.' In *Meaning and interpretation*. University of California Publications in Philosophy, 25.
- Noveck I, Chierchia G, Chevaux F, Guelminger R & Sylvestre E (2002). 'Linguistic-pragmatic factors in interpreting disjunctions.' *Thinking and Reasoning* 8, 297–326.

Formalism/Formalist Linguistics

P ten Hacken, University of Wales, Swansea, UK

© 2006 Elsevier Ltd. All rights reserved.

In formalist approaches to linguistics, the form of language is described independently of other aspects, such as its function. Mentalist approaches, including Chomskyan linguistics, Lexical-Functional Grammar, and Head-Driven Phrase Structure Grammar, aim to describe and explain the linguistic competence of a speaker. Purely formalist approaches, including Generalized Phrase Structure Grammar and Montague Grammar, study language as an abstract object. In Chomskyan linguistics, a grammar describes the individual speaker's competence, and universal grammar describes the genetically determined language faculty. The language faculty is essential for (first) language acquisition. This model is also used as a basis to explain aspects of language use.

General Characterization

Language can be studied from a variety of perspectives. Formalist approaches to linguistics have in common that they focus on the form of language. This does not mean that other aspects of language – for example, its use – are necessarily ignored. It does

mean, however, that the aim is to describe the form of language independently of these other aspects.

An alternative interpretation of the label formalist for an approach to linguistics is that the aim of such an approach is to express generalizations about language in terms of a formalism. Although this property is logically independent of the focus on form, formalist approaches generally satisfy both interpretations.

The assumption that the form of language is determined by an autonomous system is shared by all formalist approaches. Autonomy implies that the system can be described independently of other systems. It does not imply that there is no interaction with other systems. It is this autonomy that makes it worthwhile to study language through its form.

Two main positions can be identified on the question as to how this autonomous system is realized. One is the mentalist view, which assumes that language is a knowledge component in the mind/brain of the speaker. This is the position adopted, for instance, by Noam Chomsky. The other position is formalist in the narrow sense. It assumes that a language is an abstract object. This view is found especially in the logical tradition of formal semantics and implies that what individual speakers know is an imperfect reflection of the actual language.

Formalist approaches to linguistics are opposed to approaches that focus on the function or use of language. The latter include functionalist linguistics and sociolinguistics. The aim of functionalist linguistics is to explain the form of language through its function. An example of a phenomenon often studied in this way is grammaticalization. Grammaticalization is the process by which content words become function words or morphemes. Sociolinguistics studies the use of language without drawing conclusions about the nature of the underlying formal system. An example of a phenomenon studied in sociolinguistics is politeness. The relationship of these two approaches to formalist ones is not the same. Functionalist linguistics denies the autonomy of the language system, because it tries to explain it on the basis of external factors. Therefore it is not compatible with formalist approaches. Most work in sociolinguistics, however, can be interpreted as complementary to formalist approaches. Often it is neutral as to the choice of a formalist or a functionalist framework.

Among formalist approaches, Chomskyan linguistics occupies a prominent position. It is a long-established research program that attracts significant numbers of scholars. In the nearly 50 years of its existence it has provoked extensive discussion, so detailed documentation exists about its assumptions and their implications. Botha (1989) gave a systematic overview of the issues and summarized the discussion of the 1970s and 1980s. Most alternative formalist approaches to linguistics make a point of specifying how they differ from Chomskyan linguistics. For this reason, this article will first describe the research program of Chomskyan linguistics before describing the main alternative approaches within formalist linguistics. Finally, the position of language use in these approaches will be addressed.

The Research Program of Chomskyan Linguistics

Most formalist approaches to linguistics have the ambition to develop linguistics as an empirical science. In an empirical science, a theory explains the observations in the real world by describing the underlying system. Work in an empirical science is often conceived of as consisting of a cyclical succession of four stages, known as the empirical cycle. According to the empirical cycle, scientific work starts with the collection of data. The next stage is to find generalizations about these data. Then a theory is hypothesized to explain the data. Finally, this theory is tested, which supplies further data, often contradicting the theory under consideration, so that the cycle can start again. Progress in science is

the incorporation of more and more data in the explanatory scope of the theory. This model is still adopted, usually implicitly, by many researchers working in empirical science.

Logically, the empirical cycle is not sufficiently explicit to guide scientific research. At every stage, too many alternative possibilities would have to be considered. In the transition from generalizations to theories, for instance, a sufficiently imaginative researcher could come up with more theories than they could test in their lifetime. In scientific practice, however, such problems do not arise. In the perception of most researchers, it is difficult enough to come up with a single theory that fits the data. The discrepancy between the logical plethora of theories and their scarcity in scientific practice can be explained by assuming a research program. A research program is the set of assumptions guiding the researcher toward sensible decisions at any step in the empirical cycle. These assumptions need not be conscious to the researcher and are often not formulated explicitly. In the same way as a good chess player does not see all logical possible moves but only the sensible ones, a scientist working in a research program automatically constrains the possible theories to be considered.

Chomskyan linguistics is a research program in linguistics. As such, it should be distinguished from Chomsky's linguistic theory. While both were conceived by Noam Chomsky in the late 1950s, their aims and later development are strikingly different. Chomsky's linguistic theory went through a number of stages in its development, including standard theory (ST) in the 1960s, government and binding theory in the 1980s, and the minimalist program in the 1990s. In each of these stages, tree-structure representations of sentences played a role, although the constraints on the production and manipulation of these tree structures varied. Chomskyan linguistics, by contrast, remained stable during this period. It does not refer to tree structures but specifies what a linguistic theory should explain and how such a theory should be evaluated.

Chomskyan linguistics defines the object of study as the knowledge of language a speaker has. This knowledge is called the linguistic competence or internalized language (I-language). It is not open to conscious, direct introspection, but a wide range of its manifestations can be observed and used as data for the study of language.

Competence is opposed to performance. Performance includes various types of language use; for instance, a collection of utterances, a corpus of texts, a set of grammaticality judgments, and a set of results of psycholinguistic experiments. They can be used as data in linguistic research, but in every case

the relationship they have to competence should be kept in mind in their use. Naturalistic data (utterances, texts) reflect the interaction of competence with a number of other factors. Examples of such factors are knowledge of the situation in which the utterance was made, intention on behalf of the speaker, memory limitations, and degree of fatigue. Because the role of many of these factors is poorly understood, it is often difficult to use naturalistic data as evidence for the nature of competence. Experimental data also reflect competence in interaction with other factors, but in setting up the experiment, an attempt is made to control the non-competence factors. The most frequently used type of experiment is the grammaticality judgment. There is no principled reason for treating it in a special way, but because grammaticality judgments are both highly informative and easy to obtain, they have always been widely used in Chomskyan linguistics.

A grammar is a theory about the competence of an individual speaker. It describes the competence as the system underlying the observed data. As such, the grammar explains the data. The term grammar is sometimes also used to refer to competence, but this use is potentially confusing. In early stages of Chomskyan linguistics, including Standard Theory as outlined by Chomsky (1965), grammars were described in terms of rewrite rules and transformations. At this stage of the theory, rewrite rules of the type $S \rightarrow NP VP$ generated tree structures and transformations that operated on these tree structures to account for phenomena such as subject-verb agreement, passive constructions, and question formation.

The research program as described so far encounters a serious epistemological problem. If we assume a set of grammaticality judgments as our data there are infinitely many grammars (sets of rewrite rules and transformations) consistent with our data. This is a result obtained in mathematical linguistics, the theory of formal grammars. Because the aim of linguistics is to identify the grammar describing the actual competence, we need a method to distinguish this grammar from the other ones compatible with our data. Adding more data of the same type will exclude some grammars but not reduce the problem in a principled way. At any point in time, our set of data will be finite, and for any finite set of data there are indefinitely many different grammars.

As a solution to this problem, Chomskyan linguistics adopts a cross-linguistic perspective, focusing on language acquisition. In other theoretical frameworks, different languages have been compared to reconstruct a common parent language (historical-comparative linguistics) or to find universals

(typology). In Chomskyan linguistics it is assumed that certain genetically determined properties of the human mind are necessary for language acquisition. These properties are therefore reflected in one way or another in any language. By comparing different languages, linguists working in the Chomskyan framework try to find evidence of properties of the language faculty that make the acquisition of these languages possible. These properties have little in common with superficial inductive generalizations. Biologically, they are supposed to be encoded genetically in the same way as, for instance, the property that the human hand has four fingers and an opposed thumb, but they are more difficult to detect because individual languages may reflect them in different ways.

Various indications make the choice of language acquisition as the source of universals plausible. Acquisition of a language is necessary for its existence. Data from child language acquisition show that for many features of syntax, the child does not learn by repetition and resists correction. Moreover, the discrepancy between competence and performance leads to many ungrammatical and incomplete sentences in the input that children receive. The resulting I-language acquired by the child, however, is much more similar to the competence of the people in his or her environment than is a grammar that allows the sentences in the performance the child receives as input. While such considerations make a language faculty as a genetic component of the human mind plausible, conclusive empirical evidence for or against such an assumption is hard to imagine and probably impossible in principle.

In the research program of Chomskyan linguistics, the language faculty is described by a universal grammar (UG). Sometimes the language faculty is also called the language acquisition device or UG, but these names are less appropriate, because the former conflates the knowledge incorporated in the language faculty with its use and the latter conflates it with its description. The interaction of UG and grammars of individual I-languages is an essential factor in the epistemological validity of the Chomskyan research program. Since the purpose is to describe on the one hand all I-languages and on the other the language faculty that makes their acquisition possible, a proper balance has to be struck between what is genetically determined and what is acquired. If UG is too restrictive, it will not be possible to come up with grammars for all I-languages. If UG is not restrictive enough, it will not be possible to explain the learnability of the I-languages.

The interaction of the main components of Chomskyan linguistics can be represented as shown

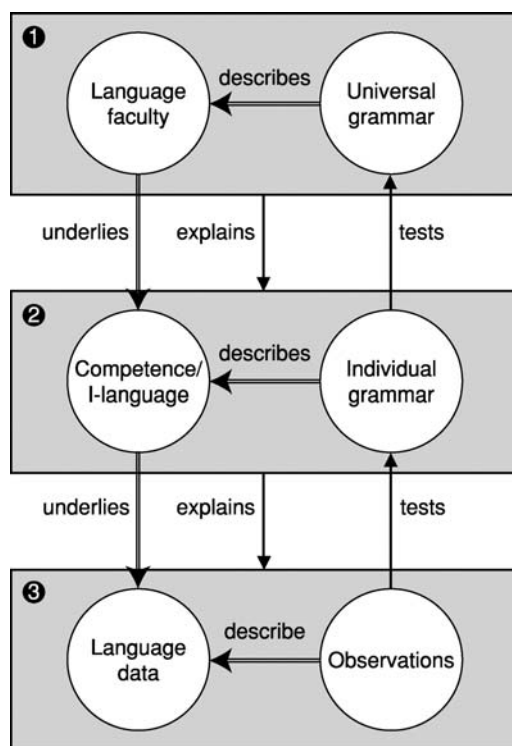


Figure 1 The research program of Chomskyan linguistics.

in **Figure 1**. Language is studied at the universal level (1), at the individual level (2), and at the level of data (3). These levels are represented as the three numbered boxes. On the left we find the real-world entities at each level: the language faculty, the individual competence, and the performance data. On the right we find the constructs of linguistics: UG, individual grammars, and observations about performance. At each level, the linguist's construct describes the real-world entity. In interpreting the relationships between the different levels, it should be kept in mind that there are many instances of the lower level corresponding to a single instance of the higher level. For example, by describing an I-language, a grammar can explain (aspects of) a large set of performance data. Conversely, these data can be used as a test for the grammar. At the higher level, UG describes the language faculty and explains thereby (aspects of) a large set of I-languages. Again, conversely, I-languages can be used as a test for UG, because UG has to allow for a descriptively adequate grammar that is in addition learnable.

While Chomsky (1965) recognized the epistemological need for UG, he also observed that ST did not incorporate the mechanisms for formulating it. By progressive generalization and cross-linguistic comparison, a model of the language faculty emerged, which was formulated as the Government

and Binding Theory (GB theory) by Chomsky (1981). In this model, UG consisted of a number of principles, which were universal. Differences between I-languages were accounted for by including parameters in the principles. A parameter specified a number of options, typically two. In language acquisition, parameters were set by selecting one of these options. The grammar of an I-language was thus determined by the parameter settings selected. The Minimalist Program (MP) outlined by Chomsky (1995) changed many of the theoretical assumptions of GB theory but continued to adhere to the principles and parameters model. An example of a difference between GB theory and the MP is the way movement was constrained. In GB theory, the individual transformation rules of ST were replaced by a single, general rule move α . This meant that anything could move anywhere unless a constraint prohibited it. In the MP, the perspective was reversed: every movement had to be motivated.

A noteworthy consequence of Chomskyan linguistics is that language can be studied at the level of individual I-languages or of the universal language faculty, but not at the level of major languages or dialects. There is no sense of **language** in Chomskyan linguistics, such that, for instance, David Beckham and Paul McCartney share the same language. Since they have different minds, they cannot share a mental component. The similarity of their I-languages, which by no means amounts to identity, can be explained as a consequence of the limited choice permitted by parameter settings, combined with a large overlap in vocabulary.

Other Mentalist Approaches

Adoption of a formalist approach to linguistics and acceptance that language is a knowledge component in the mind of the speaker by no means implies adherence to the research program of Chomskyan linguistics. There are many different ways of elaborating these two basic assumptions into a research program. Although the research program of Chomskyan linguistics has been debated more intensively and developed in more detail, this does not mean that the others are less valid in any sense. In this section, two approaches will be presented that have been developed over the past decades and continue to attract a significant group of researchers: Lexical-Functional Grammar (LFG) and Head-driven Phrase Structure Grammar (HPSG).

Lexical-functional grammar emerged in the late 1970s as a reaction against certain aspects of Chomskyan linguistics. However, the main focus of debate at the time was not the research program

of Chomskyan linguistics, as described in the previous section, but the interpretation of a number of individual theoretical results. An example of a prominent issue was the so-called psychological reality of grammars for individual languages. In any stage of Chomskyan linguistics, a sentence is represented as a tree structure generated by rewrite rules and subsequently affected by transformations that move elements of the tree to other positions. The principles and parameters of GB theory are constraints on rewrite rules and movement. Some researchers argued that if such a grammar is a psychologically real description of the speaker's competence, a sentence that requires more movement operations or more complex ones should take longer to process than one with fewer or simpler movement operations. They devised psycholinguistic tests to evaluate this hypothesis and found that processing times were not affected in the expected way. Cognitive complexity could not be related to the complexity of syntactic derivations in a transparent way. This stimulated the development of grammar formalisms in which the role of transformations was reduced or eliminated. Lexical-functional grammar is one of the most successful of them.

As the contributions to Bresnan (1982) indicated, LFG was devised mainly by Joan Bresnan. Instead of tree structures and transformations, LFG has a tree structure and a functional structure for each sentence. The functional structure represents grammatical functions, such as subject and predicate, in a feature structure. The tree structure and the functional structure are linked by a unification-based procedure. In this way, transformations are no longer required, so the problem they pose for the psychological reality of grammars no longer arises.

At first sight, the research program of LFG was not fundamentally different from that of Chomskyan linguistics. Superficially, the main difference seems to be the nature of the grammar formalisms. A grammar in LFG also intends to describe the competence of a speaker. When we consider the universal level, however, a more principled difference can be observed. Lexical-functional grammar assumes that the formal mapping procedure between the tree structure and the functional structure is universal. This assumption constrains the formulation of grammars in a way parallel to the function of UG in Chomskyan linguistics. This means that the language faculty in LFG is interpreted not as the knowledge needed for language acquisition but as a mechanism for language processing. Therefore, the research program of LFG can also be represented as in **Figure 1**, but rather than UG, a universal mechanism for processing language, i.e., mapping between form and meaning, describes the language faculty.

Another influential framework is HPSG, which emerged in the 1980s. Originally, as represented in Pollard and Sag (1987), it was strongly influenced by logical approaches to syntax (generalized phrase structure grammar; see the following section) and semantics (situation semantics). In more recent representations of the framework, Pollard and Sag (1994) and Sag and Wasow (1999) adopted a mentalist position. However, work in HPSG is more concerned with developing grammars than with discussing their status. In HPSG formalism, tree structures have been replaced entirely by feature structures. Rather than different levels of representation that are related to each other by transformations or other operations, HPSG adopts a single level. All information about what is called a sign (lexical item, phrase, sentence, etc.) is combined into a single structure. This includes phonology, syntax, semantics, and pragmatics. Relationships between components of the structure are indicated by co-indexation. This applies to both, for instance, a pronoun and its syntactic antecedent and to syntactic, semantic, and pragmatic representations of the same item. The grammar formalism in HPSG seems to be understood as a universal mechanism with language-specific parameters for building up feature structures, which suggests a research program similar to the one for LFG. The HPSG literature is in general more interested in describing individual languages than in the nature and status of universals.

Purely Formalist Approaches

Formalist approaches to grammar are often labeled as generative grammar. The term 'generative' stems from the theory of formal languages. A formal grammar is a set of rewrite rules that generates a set of sentences. The set of sentences is called a language.

In early Chomskyan linguistics, the mechanism of rewrite rules played an important role as a theoretical device. The role of the formulation of individual rewrite rules has declined in the transition to the principles and parameters model in GB theory. The conception of a language as a set of sentences goes counter to the definition of the research topic in any stage of Chomskyan linguistics. Chomsky (1986) called it an E-language (*E* for externalized), as opposed to the I-language he wanted to study.

Other linguists and philosophers pursued the idea that natural language should be studied as a set of grammatical sentences. Quine (1972) summarized the epistemological implications of this assumption. In his view, a language like English is ultimately a set of grammatical sentences. Although this set is infinite, it can still be determined for each sentence

whether it is grammatical or not. For any set of sentences, there are infinitely many possible grammars. Rather than finding the correct one among them, this approach considers all of them as equally valid. It is not the knowledge of individual people that is interesting but only the language as an abstract object. English is an abstract object, of which individual speakers have imperfect knowledge, encoded in potentially quite different, more or less imperfect grammars. This position was elaborated by Katz and Postal (1991).

In the domain of semantics, this view of language has a long tradition in the philosophical study of truth and reference. In formal logic, models were developed for calculating whether a particular logical formula is true or false, given a specific assignment of truth values to primitive statements. It was Richard Montague who extended the use of these mechanisms to complex phenomena in natural language, such as quantification. Compared to Chomskyan linguistics, Montague grammar reversed the relative importance of syntax and semantics. Whereas in Chomskyan linguistics, syntax is the central area of linguistics, Montague found syntax not interesting by itself but only as a contribution to calculating the semantics of sentences. He used categorial grammar instead of rewrite rules to represent syntax. In this formalism, constituents are characterized in terms of their relationship to referring expressions and truth values. His model-theoretic semantics was meant to characterize formally the set of possible worlds corresponding to the truth of a sentence. After Montague's death, a large number of followers continued to pursue this idea.

In the domain of syntax, the progressive formulation of principles generalizing over individual rules in Chomskyan linguistics gave rise to the objection that the grammars were no longer generative because they were not formalized in enough detail to determine whether a particular sentence would be generated or not. Many researchers who had been attracted by Chomsky's (1965) ST because of its formality did not accept this development. The introduction of these principles was motivated at least in part by the problems caused by the excessive mathematical power of transformations. Therefore, Gerald Gazdar developed a formalism that approached this problem in an alternative way and dispensed with transformations altogether. In Generalized Phrase Structure Grammar (GPSG), even long-distance dependencies are expressed in terms of a purely 'context-free grammar' (CFG). A CFG consists only of rewrite rules of the type $A \rightarrow \alpha$, where A stands for a single symbol and α for a string of symbols. An example of a long-distance dependency is found in the following: *Who*

did John think Mary loves? Here, *who* is interpreted as the object of *loves*. In Chomskyan linguistics, the usual way of expressing this is to say that *who* originates to the right of *loves* but is moved to the start of the sentence by a sequence of transformations. Gazdar demonstrated that a CFG can account for this as well as for many other relationships traditionally described with transformations. To achieve this, he decomposed syntactic categories into feature structures and formulated so-called meta-rules. A meta-rule is a rule that takes a CFG rule as input and produces another CFG rule. Meta-rules have access to individual features inside the syntactic categories.

One of the main advantages of CFGs is their easy implementation on a computer. However, Barton *et al.* (1987) demonstrated that the system of meta-rules adopted in GPSG made grammars computationally intractable. Although GPSG has lost much of its attraction and is no longer pursued as a major grammar formalism, it contributed significantly to the theory of feature structures and was at the basis of the HPSG formalism.

Formalist approaches in the narrow sense do not have a research program parallel to the ones discussed in earlier sections of this article. They do not investigate language as an empirical entity, represented in the speaker's mind but only as an abstract object. As a consequence, their research is not guided by the empirical cycle. It is more similar to formal logic or mathematics, in which theorems are derived from axioms. The axioms correspond to the grammar and the theorems to the sentences. Contrary to common practice in logic and mathematics, formal linguistics starts with a set of theorems (grammatical sentences) and non-theorems (ungrammatical sentences) and searches for a set of axioms that generates all theorems and none of the non-theorems.

Aspects of Language Use

In all formalist approaches to linguistics, the theory of the language system is intended to be part of a general, empirical account of language-related phenomena. The hypothesis that language can be described as an autonomous system, whether realized as a component of the mind/brain or as an abstract system, implies that the interaction between this system and other factors can be invoked to account for phenomena related to language use.

The use of language in communication is one of the most prominent phenomena that have to be covered in this context. Yet it should be emphasized that language and communication are independent. This contrasts with the view adopted in functionalist

approaches, where communication is invoked to explain language. In formalist approaches, language is not considered as a tool for communication (although it can be used for communication) and communication is not seen as dependent on language (although it often involves language). By recognizing this mutual independence, relevance theory, as developed by Sperber and Wilson (1986), is compatible with formalist approaches to the description of language. In HPSG, the information pertaining to language use can be encoded directly in the feature structures representing signs. In Chomskyan linguistics, it has been suggested that there is a mental component of pragmatic competence interacting with grammatical competence. Kasher (1991) elaborated this idea.

Since the 1980s, a number of other phenomena based on language use have gained a prominent position on the research agenda of Chomskyan linguistics. They include first and second language acquisition and language change. The reason for their prominence is that they can be used as a source of data about the principles and parameters involved in language.

Whereas the logical problem of language acquisition is at the basis of the language faculty hypothesis discussed in the section on Chomskyan linguistics, the practical problem of first language acquisition is analyzed in the principles and parameters model as the process of parameter setting by the child on the basis of performance data provided by people in the child's environment. The main question here is how the language faculty interacts with general cognitive development. Wexler (1999) advocated the hypothesis that the language faculty matures in the process of language acquisition. Lust (1999) proposed instead that the language faculty remains stable and that the first language competence is a distinct component from the language faculty.

Second language acquisition leads to a state of competence that is usually less than the one achieved for the first language. It is often called interlanguage. In Chomskyan linguistics, an interlanguage is an I-language. The question is what roles are played by such factors as the learner's cognitive development, the type of access to the language faculty, and the first language competence in the emergence of interlanguage competence. Gregg (1996) gave an overview of some of the main positions.

Language change is the historical development of languages such as English. In Chomskyan linguistics, English is considered an epiphenomenon. Its change means that different generations of speakers have systematic differences in language competence. Lightfoot (1999) presented a model in which it is assumed that parameter settings determining a

speaker's grammar do not change in their lifetime but their use of the grammar does. The use of certain constructions may drop below a threshold necessary for a new generation of speakers to set the parameters in the same way as their parents. When this happens, the language is perceived as having changed.

In this section, various fields have been touched upon that crucially involve language use. The role that these fields play in approaches to linguistics that assume an autonomous system of language demonstrates that this assumption does not prevent the study of language use. Instead, it guides this study in a fruitful and constructive direction. The principles and parameters model adopted in Chomskyan linguistics stimulates this type of research in particular, because it provides external evidence about the nature of principles and parameters.

See also: E-Language versus I-Language; Functional Theories of Language; Principles and Parameters Framework of Generative Grammar.

References

- Barton G E Jr, Berwick R C & Ristad E S (1987). *Computational complexity and natural language*. Cambridge, MA: MIT Press.
- Botha R P (1989). *Challenging Chomsky: the generative garden game*. Oxford: Blackwell.
- Bresnan J (ed.) (1982). *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.
- Bresnan J (2001). *Lexical-functional syntax*. Oxford: Blackwell.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky N (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky N (1986). *Knowledge of language: its nature, origin, and use*. Westport, CT: Praeger.
- Chomsky N (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Gamut L T F (1991). *Logic, language, and meaning*. Chicago: University of Chicago Press.
- Gazdar G, Klein E, Pullum G & Sag I (1985). *Generalized phrase structure grammar*. Oxford: Blackwell.
- Gregg K R (1996). 'The logical and developmental problems of second language acquisition.' In Ritchie W C & Bhatia T K (eds.) *Handbook of second language acquisition*. San Diego: Academic Press. 49–81.
- Jackendoff R S (1993). *Patterns in the mind: language and human nature*. New York: Harvester/Wheatsheaf.
- Kasher A (1991). 'Pragmatics and Chomsky's research program.' In Kasher A (ed.) *The Chomskyan turn*. Oxford: Blackwell. 122–149.
- Katz J J & Postal P M (1991). 'Realism vs. conceptualism in linguistics.' *Linguistics and Philosophy* 14, 515–554.

- Lightfoot D (1999). *The development of language: acquisition, change, and evolution*. Oxford: Blackwell.
- Lust B (1999). 'Universal grammar: the strong continuity hypothesis in first language acquisition.' In Ritchie W C & Bhatia T K (eds.), 111–155.
- Newmeyer F J (1998). *Language form and language function*. Cambridge, MA: MIT Press.
- Pollard C & Sag I A (1987). *Information-based syntax and semantics 1: Fundamentals*. Stanford, CA: Center for the Study of Language and Information.
- Pollard C & Sag I A (1994). *Head-driven phrase structure grammar*. Stanford, CA: University of Chicago Press, Chicago: Center for the Study of Language and Information.
- Quine W V (1972). 'Methodological Reflections on Current Linguistic Theory.' In Davidson D & Harman G (eds.) *Semantics of natural language*. Dordrecht: Reidel. 442–454.
- Ritchie W C & Bhatia T K (eds.) (1999). *Handbook of child language acquisition*. San Diego: Academic Press.
- Sag I A & Wasow T (1999). *Syntactic theory: a formal introduction*. Stanford, CA: Center for the Study of Language and Information.
- Sperber D & Wilson D (1986). *Relevance: communication and cognition*. Oxford: Blackwell. [2nd edn., 1995.]
- Thomason R H (ed.) (1974). *Formal philosophy: selected papers of Richard Montague*. New Haven: Yale University Press.
- Uriagereka J (1998). *Rhyme and reason: an introduction to minimalist syntax*. Cambridge, MA: MIT Press.
- Wexler K (1999). 'Maturation and growth of grammar.' In Ritchie W C & Bhatia T K (eds.), 55–109.

Frame Problem

C Viger, University of Western Ontario, London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Origins of the Problem

The frame problem, first explicitly identified by John McCarthy and Patrick Hayes (1969), arises in the attempt to design artificially intelligent machines. Intelligent reasoning includes temporal reasoning, such as planning actions or explaining a causal sequence of events. Correct temporal reasoning requires knowing the effects that actions will have in the situations in which they are performed, both the changes and the nonchanges. The frame problem, in its original formulation, is the problem for designers of artificially intelligent systems of how to represent nonchanges, such as when a cup is moved yet its color does not change. For humans such inferences are trivial, just common sense. However, designing systems with common sense has proved to be an extremely difficult problem for artificial intelligence (AI) researchers, making the frame problem of much wider interest than a technical design problem, as it touches on age-old philosophical questions about understanding intelligence – not least as it is manifest in human conversation. In the hands of philosophers, the frame problem has come to refer to a cluster of issues concerning the relevance of the content and speed of human thinking.

McCarthy and Hayes were working within the tradition of using formal logic to design intelligent systems. To handle temporal reasoning, McCarthy had developed the situation calculus, an instance of

first-order predicate logic. The formalism indexes time discretely. It includes rules that allow a system to deduce the results of an action, given a description of some situation, i.e., it can produce a description of the situation that results from the action. They applied the formalism to very simple 'toy worlds,' such as a block world consisting of only a small number of blocks on a table and a few simple actions, such as moving or stacking, that could be performed on the blocks. What McCarthy and Hayes realized was that the rules specifying the explicit effects of some action were not sufficient for the system to completely describe the situation resulting from that action. For example, if in situation S1, block A is on block B, and blocks C and D are on the table, the formalism does not allow the system to deduce that in S2, the situation resulting from moving block C onto block D, block A remains on block B. An additional rule specifying that block A remains on block B when block C is moved is required for that inference. The rules specifying the nonchanges during an action are called frame axioms, because they specify the stable frame of reference in which the action occurs.

Frame Axioms Result in Computational Overload

The difficulty with using frame axioms is not that the correct inferences cannot be drawn; it is, rather, that a very large number of frame axioms are required even in simple worlds. For example, adding a painting action to the block world requires distinct rules specifying that each block remains the same color when any block is moved or stacked or when any other

block is painted. As worlds become more complex, the number of frame axioms comes to overwhelm the system. Most of the system's computations are irrelevant deductions about what has not changed, making the system very slow and thus unable to function in real time. Also, the formalism is very brittle. Most frame axioms will be false in at least some situations when concurrent actions are possible, leading to false conclusions of the sort that humans rarely endorse. Given the complexity of our environment, it is extremely implausible that we reason using frame axioms. Thus, insofar as it is a goal of AI research to help us understand intelligent behavior, the use of frame axioms is unsatisfactory. The problem, then, is how to represent nonchanges without using a large number of frame axioms.

The Sleeping Dog Strategy and Nonmonotonic Logics

A main strategy for solving the frame problem within AI is called the sleeping dog strategy. The idea is 'to let sleeping dogs lie,' which in the context of the frame problem is to make the assumption that things remain just as they are when an action is performed unless there is some explicit indication of a change. This method uses databases containing statements describing situations. When an action is performed, its effects are computed. Statements that cease to hold are deleted, and statements describing changes are added. All other statements correspond to the nonchanges, so they are included in the database for the resulting situation; the nonchanges are stored in memory, rather than computed.

A difficulty for using the sleeping dog strategy as a solution to the frame problem is that there are often exceptional effects of actions. For example, moving a block does not ordinarily change its color, but it does in the exceptional case that the block is moved under a spray painter. Intuitively, what is required to handle exceptional cases is the ability to draw a new conclusion upon learning new information. Formally what is required is that by adding a premise, the deduction of one conclusion can be transformed into the deduction of a different, contradictory conclusion such that the original conclusion no longer follows – on pain of inconsistency. The difficulty is that classical logics do not have this property. They are monotonic, which is to say that any conclusion that follows from a set of premises also follows if further premises are added. In response, AI researchers have developed a number of nonmonotonic logics in which adding new premises blocks the derivation of the original conclusion. Using nonmonotonic logics makes it possible to derive new, contradictory

conclusions when exceptional circumstances formalized as additional premises are added to a derivation. This often involves adding a premise indicating that the action is abnormal in the given situation, so that the usual conclusion ceases to follow.

The Yale Shooting Problem

Even when nonmonotonic logics are used, there are still problems in trying to use the sleeping dog strategy. The most famous such problem is known as the Yale Shooting Problem, introduced by Hanks and McDermott (1986). The scenario in the Yale Shooting Problem is that a gun is loaded at some time t . For a number of discrete time units following t , the action taken is to wait. Then the gun is fired at Fred. The intuitively correct outcome is that in the next time unit, Fred is dead. The formalism predicts this by taking the gun's being loaded as the sleeping dog that persists. Though being alive is a state that usually persists, being shot by a loaded gun is abnormal with respect to being alive, so Fred must be dead. The formalism also predicts that Fred is alive, however, by treating the state of being alive as the sleeping dog. The reasoning in this case is that since Fred is alive, the gun must have become unloaded during the waiting, treating the action of waiting as being abnormal with respect to the gun's being loaded. The formalism cannot determine which sleeping dog, Fred's being alive or the gun's remaining loaded, should be left to rest in the given situation, so it cannot predict whether Fred is alive or dead. The problem comes about because in nonmonotonic logics the addition of a premise – the gun is shot – can undermine support for some conclusion. When more than one conclusion can become unsupported – Fred is alive, the gun is loaded – the formalism has no way to decide which one to keep, leading to contradictory predictions about the new situation. In some cases the conflict is a virtue, because it models our own uncertainty of the outcome, but even when the outcome is obvious to us, the formalism is stymied.

Holism Presents a Problem for the Sleeping Dog Strategy

A more general philosophical concern with the sleeping dog strategy results from the holistic nature of commonsense inferences. In the right situation, virtually anything we know might be relevant for successfully predicting the outcome of an action. A dropped paper clip will fall to the ground, unless it is in a magnet field. My love of pizza could even affect the New York Stock Exchange if I order the last slice of pepperoni ahead of a multibillionaire who then

invests irrationally after lunch. There is no principled way that anyone yet knows of for determining which sleeping dogs need to be woken in a particular situation. (This reading of the frame problem as a problem of accessing relevant information is of interest to linguists and philosophers of language attempting to understand what is said in making an utterance over and above its literal meaning. Research here is being conducted at the interface of pragmatics and semantics.)

AI researchers such as McDermott (1987) and Hayes (1987) criticized the broad philosophical interpretation on the grounds that it is not the real frame problem. Instead, philosophers are focused on problems of relevance and prediction, which are related to, but distinct from, the frame problem. Distinguishing these problems is more than mere pedantry; it stems from a general optimism that progress is possible following a divide-and-conquer strategy that is often successful in science. Clear distinctions also further understanding. Philosophers such as Fodor (1987, 2000) are much more pessimistic, since they see the problem as generalizing. Defining the frame problem so narrowly that it is solvable leaves the much wider problem of relevance untouched. Insofar as it is the aim of AI research to design intelligent machines, progress will be impeded until holistic reasoning can be modeled. For example, abductive reasoning, such as judging which of two hypotheses is simpler, faces the same problem of computational overload that we see in the traditional frame problem, and for the same reason, according to Fodor (2000). Whether or not some hypothesis is simple depends on the theory and background assumptions into which it is embedded. As a result, computing simplicity seems to require computing over an entire theory; yet we make simplicity judgments in real time, leaving us with the problem as to how a computational model can account for abductive inferences.

Concluding Remarks

Since the formulation of the frame problem, attempts to solve it have led to many advancements in logic and computer science that are significant in their own right, much as work on Fermat's last theorem developed mathematics. Though researchers do not agree on just what the problem is or the progress that has been made, most agree that progress on the frame problem is central to the development of cognitive science.

See also: Holism, Semantic and Epistemic; Relevance Theory; Semantics–Pragmatics Boundary.

Bibliography

- Dennett D (1984). 'Cognitive wheels: The frame problem of AI.' In Hookaway C (ed.) *Minds, machines and evolution*. Cambridge, UK: Cambridge University Press. [Reprinted in Pylyshyn (ed.), 41–64.]
- Dennett D (1996). 'Producing future by telling stories.' In Ford & Pylyshyn (eds.), 1–7.
- Dietrich E & Fields C (1996). 'The role of the frame problem in Fodor's modularity thesis: a case study of rationalist cognitive science.' In Ford & Pylyshyn (eds.), 9–24.
- Dreyfus H L & Dreyfus S E (1987). 'How to stop worrying about the frame problem even though it's computationally insoluble.' In Pylyshyn (ed.), 95–111.
- Fodor J (1987). 'Modules, frames, fridgeons, sleeping dogs, and the music of the spheres.' In Pylyshyn (ed.), 139–149.
- Fodor J (2000). *The mind doesn't work that way: the scope and limits of computational psychology*. Cambridge, MA: Bradford/MIT.
- Ford K M & Pylyshyn Z (eds.) (1996). *The robot's dilemma revisited: the frame problem in artificial intelligence*. Norwood, NJ: Ablex Publishing Corporation.
- Glymour C (1987). 'Android epistemology and the frame problem: comments on Dennett's "Cognitive wheels."' In Pylyshyn (ed.), 65–75.
- Hanks S & McDermott D (1986). 'Default reasoning, non-monotonic logics, and the frame problem.' In Kehler T, Rosenschein S, Filman R & Patel-Schneider P (eds.) *Proceedings of the Fifth National Conference on Artificial Intelligence*. Los Altos, CA: Morgan Kaufmann. 328–333.
- Haugeland J (1987). 'An overview of the frame problem.' In Pylyshyn (ed.), 77–93.
- Hayes P (1987). 'What the frame problem is and isn't.' In Pylyshyn (ed.), 123–137.
- Janlert L-E (1987). 'Modeling change – the frame problem.' In Pylyshyn (ed.), 1–40.
- Janlert L-E (1996). 'The frame problem: freedom or stability? With pictures we can have both.' In Ford & Pylyshyn (eds.), 35–48.
- Loui R P (1996). 'Back to the scene of the crime: or, who survived Yale shooting.' In Ford & Pylyshyn (eds.), 89–98.
- McCarthy J & Hayes P (1969). 'Some philosophical problems from the standpoint of artificial intelligence.' In Meltzer B & Michie D (eds.) *Machine intelligence*, vol. 4. Edinburgh: Edinburgh University Press. 463–502.
- McDermott D (1987). 'We've been framed: or, why AI is innocent of the frame problem.' In Pylyshyn (ed.), 113–122.
- Morgenstern L (1996). 'The problem with solutions to the frame problem.' In Ford & Pylyshyn (eds.), 99–133.
- Pylyshyn Z (ed.) (1987). *The robot's dilemma: the frame problem in artificial intelligence*. Norwood, NJ: Ablex Publishing Corporation.
- Shanahan M (1997). *Solving the frame problem: a mathematical investigation of the common sense law of inertia*. Cambridge, MA: The MIT Press.

Functionalist Theories of Language

C S Butler, University of Wales, Swansea, UK

© 2006 Elsevier Ltd. All rights reserved.

Functionalism within the Gamut of Linguistic Theories

Functionalist theories of language have most often been contrasted with formalist theories, particularly those developed by Chomsky during the past half century. This comparison is useful and will indeed form the starting point for the discussion in this article. However, like many dichotomies in linguistics, this one conceals a much more complex reality. A more realistic way of looking at things would be to posit a linear sequence of approaches, with strongly formalist theories such as Chomsky's at one end and the more radically functional approaches at the other. Even this, though, is still something of an oversimplification: what we really need is a multidimensional model in which the whole space occupied by linguistic theories is populated by groups of theories that form tighter or looser clusters according to their degree of similarity on a wide range of features. In what follows, we examine a number of features that are characteristic of functionalist theories. Much more detailed discussion of many of these points can be found in Butler (2003a: Chapter 1).

A further complication that should be borne in mind as we survey the range of functionalist approaches is the distinction between European and North American functionalism. Important theories initially developed in Europe include Functional Grammar (FG: Dik, 1997a, 1997b) and Systemic Functional Grammar (SFG: Halliday, 1994; Halliday and Matthiessen, 1999, 2004). In the United States, there are two rather distinct kinds of approach: on the one hand, Role and Reference Grammar (RRG: Van Valin, 2005; Van Valin and LaPolla, 1997) belongs to the rather more formal subgroup of functionalist theories; on the other hand, there is a set of approaches, more radically functionalist in nature, that have developed out of the 'West Coast' functionalism of scholars such as Givón, Haiman, Bybee, Hopper, Thompson, DuBois, Noonan, and Goldberg and that are often included, together with Langacker's cognitive grammar, under the label of 'usage-based' models. It is important to be aware that often, in the writing of American functionalists, the labels 'functionalist' and 'functionalism' are used primarily in relation to this last group of models, often with little reference to European ideas.

The Basic Tenets of Functionalism

The three most basic principles of functionalist approaches are that

- they regard communication as the primary function of language, which shapes the forms languages take
- they attribute great importance to external (cognitive and sociocultural) factors in explaining linguistic phenomena
- they reject the claim that syntax is autonomous from semantics and pragmatics.

The most basic tenet of all, from which all others derive to some degree, is that the primary purpose of language is human communication, and that this fact is crucial in explaining why languages take the form they do. This view contrasts somewhat starkly with that of Chomsky, for whom language is essentially a vehicle for expressing thought, with inter-human communication being just one of the uses to which it can be put, and not to be prioritized over other possible functions. Chomsky (1980: 229–230) defends his position by reference to what he regards as noncommunicative uses of language, such as resolving a problem or engaging in what has been called 'phatic communion' in order simply to oil the social wheels. For the functionalist, many such situations would still be communicative, though what is communicated may be largely social or interpersonal in nature rather than purely representational. Furthermore, even if some types of language use are non-communicative, they are, the functionalist will claim, merely parasitic upon the central communicative use.

If we are to study language as communication, then we will need to take into account the properties both of human communicators and of the situations in which linguistic communication occurs. Indeed, a further important claim of functionalism is that language systems are not self-contained with respect to such factors, and therefore autonomous from them, but rather are shaped by them and so cannot be properly explained except by reference to them. Linguists who make this claim belong to the group that Croft (1995) calls 'integrative functionalists,' and they undoubtedly form the largest and most influential group of functional theorists. The main language-external motivating factors are of two kinds: the biological endowment of human beings, including cognition and the functioning of language processing mechanisms, and the sociocultural contexts in which communication is deeply embedded. We might also expect that a functionalist approach would pay

serious attention to the interaction between these factors and the ways in which languages change over time, although in practice this varies considerably from one model to another.

The question of motivation for linguistic systems is, of course, not a simple one. Much of the formalist criticism of functionalist positions has assumed a rather naïve view of functional motivation, in which some linguistic phenomenon is explicable in terms of a single factor. Functionalists, however, have never seen things this way, but rather accept that there may be competing motivations, pulling in different directions and often leading to compromise solutions.

It is important to note that formalists are not necessarily totally opposed to the claim that language is functionally motivated, although this is something that is not often discussed in the formalist literature. Newmeyer (2003a), for example, claims that even hard-line formalists would accept this claim, and that Chomsky has never denied it. The question, for Newmeyer, is how much functional motivation there is, where it is located in the system, and to what extent it should form part of the research program. In particular, Newmeyer claims that functionalists and practitioners of cognitive linguistics have posited a tighter iconic form–function fit than is justifiable and is skeptical of the concept of competing motivations, on the grounds that anything can be explained away by some combination of such antagonistic forces (Newmeyer, 1998: 137–153, 2003b). The solution, of course, is to demonstrate such competing motivations and their antagonistic effects clearly so that functional ‘explanation’ does not become vacuous, a problem that is discussed in Butler (2003a: 19–25).

The great majority of functionalists also believe that within the nonautonomous linguistic system as a whole, that part of it which deals with the forms of language, and which we may call morphosyntax, is also not autonomous but is intimately bound up with the kinds of meaning that it serves to express, divided by many linguists into the semantic and the pragmatic. In other words, functionalists claim that the morphosyntax of languages is motivated by the meanings that it conveys, although we shall see that the various theories differ with regard to the strength with which they hold this position. Indeed, there are some scholars, such as Prince and Kuno, whom Croft (1995) calls ‘autonomist functionalists’ precisely because they do accept the autonomy hypothesis, but it can be argued that these approaches are peripheral to the mainstream of functionalism (Butler, 2003a: Chapter 1). Most functionalists would accept that there is some degree of (at least apparent) arbitrariness in the morphosyntax of any language, but many would again attempt

to link this to the concept of competing motivations. The functionalist position again contrasts with the formalist one, which maintains that the morphosyntax forms a system that can be described and explained independently of meaning.

Summing up the basic characteristics of functionalist theories, we may say that the core of the functionalist position is that language systems and their components are so inextricably linked with the social, cognitive, and historical contexts of language use, and with the meanings that language is used to convey, that it is futile to attempt to describe and explain them except through reference to such factors.

Further Features

There are a number of further features that, although ultimately derivable from the basic tenets, show considerable variability across functionalist theories:

- The attempt to account for the full range of linguistic phenomena rather than merely a ‘core grammar’
- The use of authentic linguistic productions as data
- Recognition and modeling of the flexibility of language in use
- Attention to discourse as well as to the sentence
- A concern with typological variation in language
- An approach to language acquisition in which the child constructs his or her language from the information available in the linguistic environment, given inherent general cognitive capacities and learning abilities

In the following sections, these points are explored in more detail.

Inclusive Rather Than Core Grammars

A truly functional approach, aiming to understand and account for language as communication, cannot restrict its scope to some kind of core grammar, as with Chomskyan theories, but must, in principle, take on the full complexity of languages and our use of them, provided, of course, that the phenomena under study are systematic.

The Use of Authentic Textual Data

If, as functionalists claim, they are interested in studying language as communication, then it is logical to assume that one of the main sources of data for their descriptions and theorizing will be samples of actual communication, in their contexts of use. In other words, we would expect functionalists to make extensive use of authentic language data in their work, particularly those large bodies of material that have been collected in the form of computer-readable corpora.

Flexibility of Meaning and Structure

If we are truly to study language in its communicative role, we must come to terms with the fact that what we say, and the way we say it, responds flexibly to the demands made by the communicative context. There are various aspects to this flexibility: the context-dependent meaning of words; the fact, clear from corpus linguistics, that a good deal of what we say is made up of reusable ‘chunks’ of language rather than generated or analyzed afresh every time we need to express a particular concept (*see*); the fuzzy nature of linguistic categories, which has been modeled through recourse to the specification of prototypes; the fact that structures in actual communication are often much less rigid than the textbook grammars would allow; the emergence of grammatical categories from the requirements of discourse; and the ‘leaky’ nature of the grammars we write to describe languages (Givón, 1995: 9).

A Discourse Grammar, Not Just a Sentence Grammar

Communication normally involves not just single, isolated sentences but also larger stretches of multi-propositional discourse. It follows that a truly functional theory should reflect this fact by containing not just a sentence grammar but also a model of discourse structure and of how the two interact.

Typological Orientation

Functionalists are interested not only in individual languages but also in ‘language’ as a whole. Since crucial motivating factors such as the human biological endowment and the overall requirements of communication are universal, we may expect that they will be reflected in linguistic universals, although it is also important to realize that because competition among motivations can be resolved in many ways, and because there are considerable differences in the sociocultural conditions under which languages are used, there are also pressures leading to diversity among languages. These concerns are manifested in the interest shown by most functionalists in linguistic typology. As pointed out by Croft (2003: 2), the label functional–typological approach is often given to an approach that seeks to motivate similarities and differences between languages in functional terms rather than in terms of the ‘formal’ explanations sought by Chomskyan linguists. Nevertheless, as Croft (2003: 5–6) also observes, the functionalist and formalist approaches share several fundamental characteristics, although with different emphases: both begin by analyzing language structures (functionalists in terms of function and Chomskians in

terms of formal properties), both are concerned with the central question of what constitutes a possible human language and posit universal constraints, both abstract patterns from the data (the functionalist across languages and the formalist within languages), and both explain universals in terms of universal human abilities (the functional typologist in terms of general cognitive and sociocultural abilities and the formalist in terms of innate principles specific to language).

A Constructivist Account of Acquisition

Functionalist theories, with their communication-in-context view of language, understandably tend to favor a constructivist view of language acquisition, according to which enough information is available in the linguistic environment for the child to be able to construct a grammar, with the help of more general cognitive factors and learning capacities. This is not to say that functionalists totally reject the possibility of innate characteristics: what is at issue is rather the nature of this endowment. For the functionalist, it is more likely that what is innate will be a set of general cognitive principles and predispositions to learn rather than some concrete set of linguistic rules and principles.

Important Functional Theories

This section examines briefly the European and American functionalist approaches mentioned at the beginning of this article, with particular reference to the features of functionalism discussed previously.

Functional Grammar

Functional Grammar was developed by the Dutch linguist Simon Dik and his colleagues from the late 1970s onwards. Dik made a strong commitment to the centrality, in the formulation of a grammar, of the concept of communicative function:

The primary aim of natural languages is the establishment of interhuman communication; other aims are either secondary or derived. (Dik, 1986: 21)

In the functional paradigm, ... a language is in the first place conceptualized as an instrument for social interaction among human beings, used with the intention of establishing communicative relationships. Within this paradigm one attempts to reveal the instrumentality of language with respect to what people do and achieve with it in social interaction. (Dik, 1997a: 3)

Dik (1986) discussed in detail the kinds of external motivation that we need to appeal to in explaining why languages are as they are, also bringing in the concept of competing motivations.

Dik (1997a: 8) also rejects the thesis that syntax is autonomous from semantics and pragmatics:

Semantics is regarded as instrumental with respect to pragmatics, and syntax as instrumental with respect to semantics. In this view there is no room for something like an 'autonomous' syntax. On the contrary, to the extent that a clear division can be made between syntax and semantics at all, syntax is there for people to be able to form complex expressions for conveying complex meanings, and such meanings are there for people to be able to communicate in subtle and differentiated ways.

He does, however, accept that grammar is a structural system, governed by a set of rules and principles, which should nevertheless be explained, wherever possible, in terms of the functioning of language as a communicative device. For example, Dik (1986: 27–28, 43) points out that expression types used to express politeness, deference, and social distance are typically longer and more complex than less polite counterparts, and that this reflects two types of iconic principle: that of diminution vs. augmentation (important = high, big; unimportant = low, small) and the avoidance of directness (see also Haiman, 1983).

The previous quotation suggests that pragmatics, as the driver of semantics, which in turn is the driving force behind the syntax, should play an important part in FG, and Dik does indeed set up a criterion of pragmatic adequacy for his grammar. He also makes a commitment to a criterion of psychological adequacy, according to which the grammar should at least be compatible with what we know of the storage and processing of language. It has been argued, however, that Dik's work failed to live up to both of these standards of adequacy. A number of new models have been put forward under the general umbrella of FG, one of whose aims is to increase the level of pragmatic and (particularly) psychological adequacy. Particularly important among these new developments are the Functional Procedural Grammar (FPG) of Nuyts, the Functional Discourse Grammar (FDG) of Hengeveld, and the Incremental Functional Grammar (IFG) proposed by Mackenzie. Mention should also be made of the Functional Lexematic Model of Mairal Usón and colleagues, which is derived from FG and Coseriu's theory of lexematics. This model pays considerable attention to cognitive aspects of lexical structuring. Brief introductions to all these models can be found in Butler (2003a, 2003b), together with evaluations of the extent to which they, as well as the parent theory of FG, attain the standards of adequacy they aspire to.

FG and its variants are committed to providing an account of the full range of phenomena found in the world's languages rather than of some

restricted core grammar. In Dik's work, most examples used to illustrate the grammar were constructed, but proponents of FG and its variants are beginning to take more seriously the need to work with authentic productions, and the use of computer-assisted corpus analysis is increasing. FG has not been among the theories that have foregrounded the flexibility of language, although there is some work on prototypes.

Dik's early model was purely a sentence grammar, although he did present a programmatic model of discourse structure in his later work (Dik, 1997b: Chapter 18). However, there have always been groups of classicists and anglicists in The Netherlands interested in the textual dimension of FG, and this interest has recently increased, with a proliferation of discourse models beginning in the 1990s. Particularly important in this connection are FDG and IFG, which aim to model discourse as well as the sentence grammar.

From its inception, FG has always had a strongly typological orientation. Dik (1997a: 14) states of FG that it should

be *typologically adequate*, i.e., that it should be capable of providing grammars for languages of any type, while at the same time accounting in a systematic way for the similarities and differences between these languages.

This orientation is very clearly reflected in the work of Dik and other FG linguists, and it persists in recent models such as FDG and IFG.

On language acquisition, Dik (1997a: 7) states the following:

From a functional point of view, ... it is certainly more attractive to study the acquisition of language as it develops in communicative interaction between the maturing child and its environment, and to attribute to genetic factors only those underlying principles which cannot be explained as acquired in this interaction.

However, only a limited amount of work on acquisition has been done so far.

Role and Reference Grammar

RRG makes a strong commitment to the study of language as communication, as shown by the following quotation from an early work in which the initial basis of the theory was developed (Foley & Van Valin, 1984: 7):

The theme underlying the various functional approaches is the belief that language must be studied in relation to its role in human communication. Language is thus viewed as a system of human communication, rather than as an infinite set of structural descriptions of sentences.

The last sentence is clearly intended to contrast the perspective taken in RRG with that of Chomskyan linguistics. With this orientation goes a concern for the functional motivation of the grammar: “RRG takes language to be a system of communicative social action, and accordingly, analyzing the communicative functions of grammatical structures plays a vital role in grammatical description and theory from this perspective” (Van Valin, 1993: 2). In practice, RRG prioritizes cognitive over sociocultural explanation, adopting what Van Valin and LaPolla (1997: 15) refer to as a ‘communication-and-cognition perspective’ and leaving out of account, for now, the ways in which language is used in different social situations (Van Valin and LaPolla, 1997: 3, 15).

The autonomy of syntax is firmly rejected (Van Valin, 1993: 2):

Syntax is not autonomous. In terms of the abstract paradigmatic and syntagmatic relations that define a structural system, RRG is concerned not only with relations of co-occurrence and combination in strictly formal terms but also with semantic and pragmatic co-occurrence and combinatory relations.

In other words, RRG postulates that semantic, pragmatic, and syntactic aspects of language, and their combinations, are all crucial in explaining the observed properties of language. Syntax, however, is only relatively motivated by semantics, pragmatics, and cognitive concerns; that is, syntactic phenomena cannot just be reduced to matters of semantics and pragmatics, and there remains some arbitrariness in the formal system.

RRG aims, in principle, to provide an account of all the grammatical phenomena attested in any language. To date, it has not made extensive use of authentic data. Like FG, it does not prioritize the concept of flexibility in its account of languages. It does, however, make use of the concept of prototype, especially in its use of hierarchies of argument types to predict more and less typical choices in two areas: (1) the ‘macroroles’ of Actor and Undergoer in the clause, which are generalizations over more specific thematic relations, and (2) syntactic relations, which RRG conceptualizes not in terms of the usual functions of Subject and Object but, rather, in terms of the notion of the ‘privileged syntactic argument’ of a construction. Although RRG does offer a detailed account of a number of discourse pragmatic areas, such as information distribution (in terms of topic and focus) and switch reference, it has not developed its own model of discourse structure and is still essentially a sentence grammar.

Like FG, RRG is very much concerned with typological matters: “The theory is greatly concerned with

typological issues. In particular, it seeks to uncover those facets of clause structure which are found in all human languages” (Van Valin, 1993: 4). Indeed, Van Valin says that the theory arose out of the question of what linguistic theory might look like if, rather than being centered on English, it were based on languages such as Dyirbal, Lakhota, or Tagalog.

RRG takes a constructivist line on language acquisition (Van Valin, 1993: 2):

The RRG approach to language acquisition ... rejects the position that grammar is radically arbitrary and hence unlearnable, and maintains that it is relatively motivated (in Saussure’s sense) semantically and pragmatically. Accordingly, there is sufficient information available to the child in the speech to which it is exposed to enable it to construct a grammar, and therefore the kinds of autonomous linguistic structures posited by Chomsky are unnecessary.

There have been a number of studies using RRG as a basis for accounting for observed patterns in the acquisition of morphosyntax in a variety of languages (for a summary, together with a discussion of the relationship between acquisition studies and the linguistic theory, see Butler, 2003b: 402–413).

Systemic Functional Grammar

SFG, as developed by Halliday, has its roots in the work of the British linguist J. R. Firth, with influence also from Hjelmlev, the anthropologist Malinowski, and the Prague School. Halliday makes it clear that his aim is to give an account of linguistic communication, or ‘language in use,’ and that languages are as they are because of the functions they have evolved to serve:

Language has evolved to satisfy human needs; and the way it is organized is functional with respect to these needs – it is not arbitrary. A functional grammar is essentially a ‘natural’ grammar, in the sense that everything in it can be explained, ultimately, by reference to how language is used. (Halliday, 1994: xiii)

Both the general kinds of grammatical pattern that have evolved in language, and the specific manifestations of each kind, bear a natural relation to the meanings they have evolved to express. (Halliday, 1994: xviii)

Note that Halliday takes a very strong stand on functional motivation: in principle, everything in the grammar can be motivated through function – a position that some other functionalists would regard as rather extreme. As might be expected from the antecedents of SFG in the work of Firth and Malinowski, Halliday has always been concerned more with social and cultural motivations than with those related to psychological/cognitive processes: indeed, the social dimension is more strongly developed in SFG than in

any other functional theory. This is particularly apparent in the theory of social context developed by Halliday and colleagues in the 1960s and since developed not only by Halliday but also by Hasan, Martin, and others. SFG distinguishes between dialectal and diatypic (or ‘register’) variation in language, and within the latter between the variables of field (concerned with the nature of the social action), tenor (involving the relationships between interactants in communication), and mode (the medium of communication and its role in the linguistic event). SFG has also developed accounts of genres, seen as ways of getting things done linguistically.

SFG totally rejects the concept of an autonomous syntax. The theory is firmly centered on meaning, with syntactic patterns being seen as one mechanism for realizing meanings, a means to an end. Indeed, Halliday prefers to avoid the term ‘syntax’ altogether in order to distance himself from the view that syntax can be opposed to semantics and pragmatics. Instead, he uses the term ‘lexicogrammar,’ chosen to emphasize the claim that grammar and lexis are not totally different kinds of patterning but are part of the same level, different in degree rather than in kind. Both act, together with intonation, to give shape to the meanings we wish to convey. Furthermore, the division between lexicogrammar and semantics is itself not a clear one: “There is no clear line between semantics and grammar, and a functional grammar is one that is pushed in the direction of the semantics” (Halliday, 1994: xix). SFG has not, historically, given high priority to typological matters, and this is reflected in the fact that the most comprehensive accounts of this theory (Halliday, 1994, 2004; Halliday and Matthiessen, 1999) are centered very firmly on English. However, currently there are signs of greater interest in typological matters (Caffarel *et al.*, 2004). The approach to typology in SFG differs sharply from that in FG and RRG in that the emphasis is on the detailed description of particular languages, followed by generalization where possible, rather than on using data from a wide range of languages in order to formulate the theory in the first place.

SFG takes a strongly constructivist line on language acquisition, and there have been longitudinal studies of three individual English-speaking children (for summaries, discussion, and references, see Butler, 2003b: 413–424).

We should note here the existence of a second version of SFG, developed by Fawcett and colleagues. This model takes Halliday’s work as its starting point but modifies it in a number of important ways, one of which is to propose clearly separate levels of semantics and syntax (Fawcett, 2000). However,

this version still treats the semantics as the underlying driving force at the center of the model, with syntax as one means (together with lexical items and intonation) for the realization of meaning.

West Coast Functionalism

So-called West Coast functional grammar was never a single, unified model but, rather, a fairly loose collection of individual approaches. We cannot, therefore, survey the whole area here but will concentrate on one important exemplar, the work of Givón.

Givón subscribes fully to the central tenets of functionalism discussed previously. He (1995: xv) holds that “language (and grammar) can be neither described nor explained adequately as an autonomous system,” but that in order to understand why language is as it is, we must make reference to “the natural parameters that shape language and grammar: cognition and communication, the brain and language processing, social interaction and culture, change and variation, acquisition and evolution.” It is, however, possible for competing motivations to lead to a situation in which grammar “assumes its own reality” (Givón, 1995: 11). The central part played by semantics, pragmatics, and discourse within Givón’s approach to language is abundantly clear in all his work. Furthermore, relationships between linguistic usage, synchronic variation, and language change are discussed in terms of the grammaticalization of originally lexical elements.

Givón rejects Chomsky’s sanitization of the data for linguistic theory building, through the prioritization of competence over performance. He thus takes an inclusive stance on the range of phenomena for linguistic study, and makes frequent use of attested linguistic data, especially in his studies of the frequency distribution of features in different languages and their varieties. Although Givón does not offer a model of discourse structure as such, his work is very much involved with the relationship between grammar and discourse, and context is a key concept in his work. The flexibility of language is an important theme in Givón’s writing, and the concept of prototype is central.

Givón’s work is strongly typological in nature, striking a balance between what is common across languages and their individual characteristics. He (1995: 17) comments that “[w]hile recognizing cross-language typological diversity, most functionalists remain committed to the existence of language universals, presumably both of meaning/function and grammatical structure.”

Finally, Givón takes an explicitly constructivist position on language acquisition, claiming that what the child initially acquires is not syntax, in the

Chomskyan sense, but rather a more rudimentary communicative system, which only later becomes modified toward syntax. He castigates Chomsky for his rejection of communicative function, general cognitive capacities, and sociocultural effects as important factors in acquisition.

Usage-Based Functionalist–Cognitive Models

The label usage-based is used for a group of models that claim that “the speaker’s linguistic system is fundamentally grounded in ‘usage events’: instances of a speaker’s producing and understanding language” (Kemmer and Barlow, 2000: viii). This label brings together models derived from earlier work in West Coast functionalism with approaches such as Langacker’s cognitive grammar and some variants of construction grammar, such as Goldberg’s model. Here, we concentrate on those approaches that derive from the earlier work of West Coast functionalists.

One important strand of usage-based functional linguistics is Hopper’s Emergent Grammar, which takes a very radically functionalist stand in claiming that “[g]rammar . . . is not the source of understanding and communication, but a by-product of it. Grammar is, in other words, epiphenomenal” (Hopper, 1998: 156). Thus, a grammar is seen not as a unified, relatively stable system but, rather, as “an open-ended collection of forms that are constantly being restructured and resemanticized during actual use” (Hopper, 1998: 159). This model is thus firmly rooted in the communicative functions of language and in the cognitive and social contexts of language use. It also pays great attention to relationships between usage, synchronic variation, and diachronic change, through the study of grammaticalization.

Closely related to emergentism is work by Thompson, Hopper, Bybee, and others, which prioritizes the study of authentic conversational data, and in which the concept of broad syntactic templates is replaced by local, lexically bound schemas or collocations, often formulaic in nature, which are specific to the language concerned. These reusable fragments, often with slots for variable items, are seen as empirically verifiable instantiations of the concept of construction, seen as a pairing of form and meaning/function, which is central to construction grammar models. For instance, Thompson (2002) presents evidence that in conversational English what are traditionally analyzed as complement-taking predicates (e.g., *think* + complement clause) are better seen in terms of the combination of (1) a fragment, often formulaic in nature, with epistemic, evidential, or evaluative function, such as *I think/I don’t think/I thought/I*

didn’t think/I guess/I remember, etc., and (ii) a declarative or interrogative clause that carries the main representational content.

Clearly, these usage-based approaches adopt a very inclusive stance on the range of phenomena that are to be accounted for, since the basic aim is to provide an explanatory treatment of the rich diversity to be found in authentic interaction. These models are very much concerned with the flexible responses of language to the contexts in which it is used: they are rooted in the discourse and the ways in which it conditions the choice of what is said and how it is expressed. Considerable weight is given to the frequency of particular items and combinations in language use, on the grounds that frequency demonstrably affects a range of phenomena, including phonological reduction, constituent structure, and the retaining of irregular or even archaic patterns (Bybee and Hopper, 2001). Typological data are also crucial and are taken to support the paramount importance given to discourse motivations for grammatical structuring. Finally, a clearly constructivist line is taken with respect to language acquisition (Tomasello, 2003).

Conclusions

Functionalism, like other terms for ‘schools’ of linguistics, is a convenient label for a complex, varied set of approaches to linguistic theory and description. We have seen, however, that these approaches are united by rejection of the claim that the linguistic system should be studied independently of the cognitive, sociocultural, and temporal factors that at least partially motivate it and also by rejection of the claim that syntax is autonomous from semantics and pragmatics. Functional models regard language as primarily a means for human communication in context and attempt to explain as much as possible in terms of functional motivations, which may compete to give the appearance of arbitrariness in the system. Many also exhibit a number of derived characteristics, although these are much more variable across models: They are not confined to a core grammar but, rather, attempt to take on board the full range of systematic linguistic phenomena; they often use authentic linguistic productions as part of their database; they attempt to model the flexibility and ‘fuzziness’ of language; they are concerned with discourse as well as the sentence grammar; they study typological variation in language; and they adopt a constructivist approach to the acquisition of language by the child.

Finally, it is important to note that there are important similarities between functional models, as defined

here, and explicitly cognitively oriented approaches such as Cognitive Grammar and the various models of Construction Grammar. This is reflected, for example, in the fact that recent usage-based models embrace both functional and cognitive orientations (see also Horie and Comrie, 2000).

See also: Principles and Parameters Framework of Generative Grammar.

Bibliography

- Butler C S (2003a). *Structure and function: a guide to three major structural-functional theories. Part 1: approaches to the simplex clause*. Amsterdam: Benjamins.
- Butler C S (2003b). *Structure and function: a guide to three major structural-functional theories. Part 2: from clause to discourse and beyond*. Amsterdam: Benjamins.
- Bybee J & Hopper P J (2001). *Frequency and the emergence of linguistic structure*. Typological studies in language No. 45. Amsterdam: Benjamins.
- Caffarel A, Martin J R & Matthiessen C M I M (eds.) (2004). *Language typology: a functional perspective*. Amsterdam: Benjamins.
- Chomsky N (1980). *Rules and representations*. New York: Columbia University Press.
- Croft W (1995). 'Autonomy and functionalist linguistics.' *Language* 71, 490–532.
- Croft W (2003). *Typology and universals*. Cambridge textbooks in linguistics (2nd edn.). Cambridge, UK: Cambridge University Press.
- Dik S C (1986). 'On the notion "functional explanation."' *Belgian Journal of Linguistics* 1, 11–52.
- Dik S C (1997a). *The theory of Functional Grammar, part 1: the structure of the clause* (2nd edn.). Hengeveld K (ed.) Functional Grammar series No. 20. Berlin: de Gruyter.
- Dik S C (1997b). *The theory of Functional Grammar, part 2: complex and derived constructions*. Hengeveld K (ed.) Functional Grammar series No. 21. Berlin: de Gruyter.
- Fawcett R P (2000). *A theory of syntax for Systemic Functional Linguistics*. Current issues in linguistic theory No. 206. Amsterdam: Benjamins.
- Foley W A & VanValin R D Jr (1984). *Functional syntax and universal grammar*. Cambridge, UK: Cambridge University Press.
- Givón T (1995). *Functionalism and grammar*. Amsterdam: Benjamins.
- Haiman J (1983). 'Iconic and economic motivation.' *Language* 59, 781–819.
- Halliday M A K (1994). *An introduction to functional grammar* (2nd edn.). London: Arnold.
- Halliday M A K (2004). *An introduction to Functional Grammar* (3rd edn., revised by C M I M Matthiessen). London: Arnold.
- Halliday M A K & Matthiessen C M I M (1999). *Constructing experience through meaning: a language-based approach to cognition*. London: Cassell.
- Hopper P J (1998). 'Emergent grammar.' In Tomasello M (ed.) *The new psychology of language: cognitive and functional approaches to language structure*. Mahwah, NJ: Erlbaum. 155–175.
- Horie K & Comrie B (2000). 'Introduction.' In Horie K (ed.) *Complementation (cognitive and functional perspectives)*. Amsterdam: Benjamins. 1–10.
- Kemmer S & Barlow M (2000). 'Introduction: a usage-based conception of language.' In Barlow M & Kemmer S (eds.) *Usage-based models of language*. Stanford, CA: CSLI. vii–xxviii.
- Newmeyer F J (1998). *Language form and function*. Cambridge, MA: MIT Press.
- Newmeyer F J (2003a). 'Grammar is grammar and usage is usage.' *Language* 79(4), 682–707.
- Newmeyer F J (2003b). 'Meaning, function, and the autonomy of syntax.' In Moore J & Polinsky M (eds.) *The nature of explanation in linguistic theory*. Stanford, CA: CSLI. 263–280.
- Thompson S A (2002). '"Object complements" and conversation: towards a realistic account.' *Studies in Language* 26, 125–163.
- Tomasello M (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- VanValin R D Jr (1993). 'A synopsis of Role and Reference Grammar.' In Van Valin R D Jr (ed.) *Advances in Role and Reference Grammar*. Amsterdam: Benjamins.
- VanValin R D Jr (2005). *Exploring the syntax-semantics interface*. Cambridge, UK: Cambridge University Press.
- VanValin R D Jr & LaPolla R J (1997). *Syntax: structure, meaning and function*. Cambridge, UK: Cambridge University Press.

Future Tense and Future Time Reference

Ö Dahl, Stockholm University, Stockholm, Sweden

© 2006 Elsevier Ltd. All rights reserved.

It is tempting to think of time simply as a line extending in both directions from the point at which we happen to be located. However, in constructing a theory of temporal semantics, we have to acknowledge that what is ahead of us – the future – is epistemologically radically different from both what is behind us – the past – and what is taking place at this moment – the present. Future states of affairs cannot be perceived or remembered, although they can be the subject of our hopes, plans, conjectures, and predictions. Philosophers such as Aristotle have claimed that the future has a special nature not only epistemologically but also ontologically: statements about the future do not yet have a determinate truth value. In a possible worlds framework, the ‘branching futures model’ can be seen as an expression of a similar idea: time looks like a tree rather than a line, and at any point in the tree there is only one way back into the past, but many branches lead into the future.

Against this background, it is perhaps not so strange that there tend to be asymmetries in the ways in which temporal reference is structured in languages, and that in particular the grammatical category of tense often blurs into modality and evidentiality in the area of the future. Whether for instance the English auxiliaries *shall* and *will* should be seen as markers of future tense or rather as ordinary modal verbs is a much-debated issue, the importance of which depends on the stand one takes on another, equally contentious, issue: how essential it is to uphold the discreteness of grammatical categories. If it is acknowledged that it is normal for the semantics of grammatical items to combine temporal elements with components of a modal, evidential, or aspectual character, it may become more important to study how the weight of these different factors shift over time, in the process of grammaticalization. From this perspective, it is notable that the diachronic sources of what grammars refer to as future tenses typically have exclusively nontemporal meanings, and the temporal meaning elements tend to grow stronger during the course of grammaticalization (‘temporalization,’ in the term of Fleischman, 1983), as future markers gradually obtain an obligatory status. English is a language that has advanced relatively far along the road towards obligatory marking of future time reference. In this regard, it is instructive to compare

English to a language such as Finnish, in which there is hardly any grammaticalization of future time reference. In English, the sentence *It is cold tomorrow*, with the present tense of the copula *is*, sounds rather strange: it is more natural to say *it will (it’ll) be cold tomorrow* or *it is going to be cold tomorrow*. In Finnish, on the other hand, we may replace the adverb *tänään* ‘today’ in the sentence *Tänään on kylmää* ‘Today is cold’ with *huomenna* ‘tomorrow,’ yielding *Huomenna on kylmää* ‘(lit.) Tomorrow is cold,’ without any further changes in the sentence. Thus, Finnish weather forecasts are typically formulated in the present tense, which is hardly possible in English.

The English examples, however, also illustrate two other important points. First, the obligatoriness of future marking in English is not independent of the epistemological status of the statement. If it concerns an event that is fixed by some kind of schedule, English tends to use the present tense, although the time referred to is in the future, as in *The train leaves at noon*. Second, future marking is an area where we often find competition between two or more grammatical devices. For English, *will* and *be going to* have already been mentioned, but there are in fact several additional ways of referring to the future, such as by the present progressive (*We are leaving at four*) or by a combination of *will* and the progressive (*The shop will be closing in five minutes*), neither of which have a progressive meaning in the examples cited. Other languages are similar: Bybee *et al.* (1994) found at least two futures in 70% of the languages in their sample and at least three in close to 25%. As the word ‘competition’ suggests, the choice between the alternative ways of marking future is usually not reducible to any simple semantic or pragmatic distinction: rather, a number of different factors are at play: thus, *will* and *be going to* differ both semantically and stylistically. In many cases, differences between future-marking devices are attributable to what point they have reached in the grammaticalization process; in others, the differences reflect the original meanings of the diachronic sources of the items in question.

Future-marking devices derive historically from a number of sources. Among the most common are auxiliary constructions expressing obligation (‘must’), e.g., English *shall*; volition/intention (‘want’), e.g., English *will*; and motion (‘go’ and ‘come’), e.g., English *be going to*. However, a future tense may develop out of an earlier nonpast or imperfective as an indirect effect, for example of the expansion of an earlier

progressive – the future uses are what is left of the old category after that expansion (possible examples mentioned by Bybee *et al.* (1994) are Margi, Tigre, Pangasinan, and Kui).

In the development of futures, several things happen. To start with, there is normally an initial change of meaning that may involve both what has been called ‘pragmatic strengthening’ and ‘semantic bleaching.’ Thus, a verb of volition, such as *want*, does not normally imply that the willed action is performed; to interpret something like *She wants to leave* as ‘She will leave,’ the meaning of the volitional verb has to be strengthened. But when extended to cases of ‘pure prediction’ such as *It will rain*, the volitional element has to be bleached altogether. Furthermore, the item gradually comes to be used in contexts where it is communicatively redundant, which leads to it being reduced phonetically (e.g., *is going to* > *is gonna*). Eventually, the grammaticalizing item may fuse with the main verb and become affixed to it. A famous example is the French inflectional future as in *il chantera* ‘he will sing,’ which derives from a construction involving the Latin verb *habere* ‘have’ in an obligative meaning. English *shall/will* have not become inflections, although they are usually cliticized to the preceding word in the reduced form *’ll*.

Being usually more advanced on the grammaticalization path, inflectional futures tend to have a wider range of uses than periphrastic ones. Normally, future-marking devices start out in main clauses, which are bearers of the illocutionary force of an utterance. English is an example of a language in which no future time marking is normally found in some types of subordinate clauses, such as conditionals and temporal clauses – for example, *If it rains, you’ll get wet*, where only the main clause is marked for future time reference. As it turns out, future marking in such clauses tends to be restricted to languages with inflectional futures (Bybee and Dahl, 1989).

It was said above that future states of affairs are the subjects of our hopes, plans, conjectures, and predictions. The latter notions also represent what we can call different epistemological bases for statements about the future, and, as was also mentioned above, the way in which such statements are expressed in a language can depend on which kind of epistemological base it has. A major distinction may be drawn between **intention-based** and **prediction-based** future time reference. Particularly in everyday conversation, a large part of what is said about the future refers to plans and intentions of the participants. I announce what I intend to do, or

ask you what you intend to do. This is clearly different from discussing what the weather will be like tomorrow. A straightforward grammatical opposition based on the distinction between intention-based and prediction-based future time reference is less common than one would perhaps think, in view of the apparent cognitive salience of that distinction. Its importance lies rather in the observation that markers that are originally restricted to intention-based future time reference tend to develop into general future markers, which include prediction-based future time reference as central cases but can in the normal case still be used for intention-based future time reference.

Another major parameter is the temporal distance between the speech act and the future time point in the future. Immediacy is often cited as a condition on the use of certain future-marking devices, such as the English *be going to* or the French *aller* + infinitive construction. At a closer look, it often turns out that immediacy is a contributing factor but hardly the only one, as we shall illustrate below. It does happen, however, that more precise restrictions on temporal distance develop, although this is much less common for the future than for the past. Bybee *et al.* (1994) cite Mwera (1994; original source Harries, 1950) as an example of a language that has three different future auxiliaries, *ci* for reference to the same day (hodiernal future), *cika* for the following day (crastinal future), and *jiya*, which they interpret as a general future.

Typological surveys (Dahl, 1985; Bybee, 1994) have shown approximately equal numbers of inflectional and periphrastically expressed futures. In the sample in Dahl and Velupillai (2005), North America, Australia, central New Guinea, the Caucasus, and South Asia come out as areas where languages with inflectional futures are in a clear majority. Among areas where inflectional futures tend to be absent are Southeast Asia (where most languages are isolating) and northern Europe.

See also: Modal Logic; Temporal Logic; Tense and Time: Philosophical Aspects.

Bibliography

- Bybee J & Dahl Ö (1989). ‘The creation of tense and aspect systems in the languages of the world.’ *Studies in Language* 13, 51–103.
- Bybee J, Perkins R & Pagliuca W (1994). *The evolution of grammar: tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

- Dahl Ö (1985). *Tense and aspect systems*. Oxford: Blackwell.
- Dahl Ö & Velupillai V (2005). 'Grammatical marking of future tense.' In Comrie B, Dryer M, Gil D & Haspelmath M (eds.) *World atlas of linguistic structures*. To be published by Oxford University Press.
- Fleischman S (1983). 'From pragmatics to grammar: diachronic reflections on complex pasts and futures in Romance.' *Lingua* 60, 183–214.
- Harries L (1950). *A grammar of Mwera*. Johannesburg: Witwatersrand University.
-



Game-Theoretical Semantics

J Hintikka and G Sandu, University of Helsinki,
Helsinki, Finland

© 2006 Elsevier Ltd. All rights reserved.

The leading ideas of game-theoretical semantics (GTS) can be seen by considering the truth condition of a sentence S of an interpreted first-order language. Now, S is true in an obvious pretheoretical sense iff there exist suitable witness individuals testifying to its truth. Thus, $(\exists x)F[x]$ is true iff there exists an individual b such that $F[b]$, $(\forall x)(\exists y)F[x,y]$ is true iff for each a there exists an individual b such that $F[a,b]$, and so on. As such examples show, one witness individual can depend on others. Hence, the existence of suitable witness individuals for S means the existence of a full array of Skolem functions for S , such as the function f in the sentence $(\exists f)(\forall x)F[x,f(x)]$, which is equivalent to $(\forall x)(\exists y)F[x,y]$. Such arrays of Skolem functions can be seen to codify a winning strategy in a simple game (semantical game) between a ‘verifier’ V and a ‘falsifier’ F . The game $G(S)$ associated with S is played on some domain in which the nonlogical constants of S have been interpreted. $G(S)$ begins with S , but at each move the sentence that the players are considering changes. When a move is made, V chooses disjuncts and values of existential quantifiers, whereas F chooses conjuncts and values of universal quantifiers, proceeding from outside in. For $G(\sim S)$ a game rule tells V and F to exchange roles and continue as for $G(S)$. If a play of $G(S)$ ends with a true negated or unnegated atomic sentence or identity, V wins and F loses; if it ends with a false one, vice versa. S is true iff there exists a winning strategy for V and false iff there exists one for F . GTS amounts to the systematic use of such truth conditions.

Game-theoretical ideas thus enter into GTS as highlighting the role of Skolem functions in logical theory. For this purpose, the whole generality of game theory is usually not needed. For instance, the only failure of perfect information that needs to be considered is a player’s ignorance of earlier moves because only the individuals chosen at such moves can be arguments of Skolem functions.

So far, GTS for quantified sentences merely spells out our ordinary concept of truth. In traditional first-order languages, truth in the sense of GTS coincides with truth according to a Tarski-type definition (assuming the Axiom of Choice). But the GTS treatment can be extended and varied in several ways not available otherwise.

1. By allowing informational independence in the sense of game theory, we obtain a more expressive logic called independence-friendly (IF) logic (cf. Hintikka and Sandu, 1989). The greater expressive power is due to the semantic job of quantifiers as expressing dependence relations between actual variables by means of their formal dependence on one another. In the received first-order logic, only some patterns of such dependence could be formulated because the dependence-indicating scope relation is nested and hence incapable of expressing other dependence patterns. This defect is eliminated in IF logic. The role of Skolem functions in GTS is illustrated by the fact that each IF first-order sentence has a sigma-one-one second-order translation, namely, the sentence that asserts the existence of its Skolem functions. Conversely, each sigma-one-one sentence can be translated into the corresponding IF first-order language.

2. The independence of a quantifier (Q_2y) from another quantifier (Q_1x) within whose syntactical scope it occurs can be expressed by writing it (Q_2y/Q_1x). Often it is convenient to restrict this notation to quantifiers of the form $(\exists y/\forall x)$ and to assume that existential quantifiers are independent of one another, as are universal ones. (This simplification is used in this article.) As this slash notation suggests, the semantics of unextended first-order IF languages is not compositional. The limits of compositional methods in the semantics of IF logics have been investigated intensively (Hodges, 1997), and the impossibility of compositional semantics for IF logic normally interpreted has been shown (Cameron and Hodges, 2001; Sandu and Hintikka, 2001). The strength and naturalness of IF logic thus throws serious doubts on compositionality as a desideratum in semantic

theorizing in general. In any case, many noncompositional languages can be treated semantically by means of GTS.

3. GTS can be extended to natural languages by allowing the substitution of names for entire quantifier phrases, as in Hintikka and Kulas (1983, 1985). Then the meaning of this phrase has to be taken into account by suitable additions to the output sentence. For instance, a game step (amounting to existential instantiation) might take the players from a singular sentence of the form

(1) $X - \text{some } Y \text{ who } Z - W$

to a sentence of the form

(2) $X - b - W, b \text{ is a } Y \text{ and } b \text{ Zs}$

Perhaps the greatest difference between natural language games and formal ones is that in the former the order of the application of game rules can be indicated by means other than scopes (labeled tree structures), for instance by the lexical items in question. This shows the limitations of the notion of scope, as argued in Hintikka (1997). At the very least, the two functions of scope (parentheses) have to be distinguished from one another as indicating the limits of binding (binding scope) and as indicating logical priority (priority scope). This distinction helps, among other things, to solve the problem of donkey sentences.

4. GTS can accommodate any concept, logical or nonlogical, whose meaning can be captured by a rule or rules in a semantical game. Cases in point are anaphoric pronouns, epistemic notions, genitives, *only*, and so on (cf. Hintikka and Sandu, 1991). For instance, the semantics of anaphoric pronouns can be captured by construing them, in effect, as existential quantifiers ranging over the individuals hitherto selected by V and F in a play of a semantical game. This extendability of GTS to nonlogical concepts throws into doubt the possibility of any sharp distinction between logical and nonlogical concepts.

By means of IF logic, several mathematical concepts can be expressed on the first-order level that could not be captured in ordinary first-order logic, including equicardinality, infinity, (topological) continuity, and König's lemma. In general, IF logic extends greatly the scope of what can be done in mathematics on the first-order level (cf. Hintikka, 1996).

5. The notion of informational independence plays an especially important role in epistemic logic, including the logic of questions and answers. Their logic depends essentially on the logical properties of the desideratum of a question. Such a desideratum is of the form K_1S where K_1 means *I know that*.

It expresses the epistemic state that an answer to the question is calculated to bring about. The question ingredient is now of the form $(\exists x/K_1)$ for wh-questions and (\forall/K_1) for propositional questions. All the most important notions related to questions and answers can be defined by means of the slash notation.

6. The law of the excluded middle amounts to the requirement of determinacy in semantical games and, hence, is not always satisfied. Determinacy fails, in fact, in IF logic, where the negation \sim is used but not the contradictory \neg negation. The latter can be introduced by a fiat, but within the GTS framework this can be done only when it occurs sentence-initially. When both negations are present, we obtain a logic whose algebraic structure is that of a Boolean algebra with an operator in Tarski's sense. In that logic, we can define generalizations of such notions as orthogonality and dimension.

7. The failure of *tertium non datur* shows that IF logic is closely related to intuitionistic logic. In a sense, IF logic is in fact more elementary than ordinary first-order logic. For instance, an application of the GTS truth definition to a sentence does not involve infinite closed totalities of individuals even when the domain is infinite. If such totalities are relied on, we can give truth conditions also for sentences in which \neg occurs internally. The resulting logic is as strong as the entire second-order logic (with the standard interpretation), even though it is first-order logic in the sense that all quantifiers range over individuals.

In general, GTS shows that our concept of negation is irreducibly ambiguous between the strong (dual) negation \sim and the contradictory negation \neg . This fact explains a number of features of the behavior of negation in natural language (cf. Hintikka, 2002a).

8. The simplest type of IF sentence that is not translatable into ordinary quantificational notation is of the following form, known as a Henkin quantifier sentence:

(3) $(\forall x) (\forall y) (\exists z/\forall y) (\exists u/\forall x) F[x, y, z, u]$

The negation $\sim S$ of an IF sentence S can be formed in the same way as in ordinary logic. For instance, the negation of (3) is (4)

(4) $(\exists x) (\exists y) (\forall z/\exists y) (\forall u/\exists x) \sim F[x, y, z, u]$

The 'independence-free' meaning of (4) is brought out by its equivalence with (5)

(5) $(\exists f) (\exists g) (\exists x) (\exists y) (\forall z) (\forall u) \\ ((z = f(x) \ \& \ u = g(y)) \supset \sim F[x, y, z, u])$

9. IF logic is misnamed in that it allows for the representation of more dependencies (not just more

independencies) than ordinary logic. A case in point is constituted by irreducibly mutually dependent variables. In order to express them, we have to generalize the notion of function and to take seriously the idea that a functional identity $y = f(x)$ expresses a dependence relation (cf. Hintikka, 2002b). This can be taken to mean that if such an identity fails to have a truth value for some x_o , $f(x_o)$ has to be taken to represent a probability distribution. Such distributions must then be able to be reified so as to act as argument values, too. With this understanding, the mutual dependence of x and y satisfying the condition $F[x, y]$ can be expressed by

$$(6) (\exists f) (\exists g) (\forall x) (\forall y) (x = f(y) \ \& \ y = g(x) \ \& \ F[x, y])$$

which is obviously equivalent to

$$(7) (\forall x) (\forall y) (\exists z/\forall x) (\exists u/\forall y) (x = z \ \& \ y = u \ \& \ F[x, y])$$

10. The GTS approach can be varied in other ways, for instance, by restricting V's strategies to recursive or otherwise constructive ones as discussed in Hintikka (1996). In a different direction, a semantic game can be divided into subgames with specific rules for the transfer of information from one subgame to another one. Such subgames can be used for the interpretation of conditional sentences (cf. Hintikka and Sandu, 1991).

11. The notion of truth is put into a new light by IF logic. Tarski's well-known result shows that truth cannot be defined in an ordinary first-order language L for L itself, even when the syntax of L can be represented in L, for instance by means of Gödel numbering. The reason is that quantifiers ranging over numbers as numbers and quantifiers ranging over numbers as codifying numerical expressions must be informationally independent of one another. This requirement cannot be implemented in the received first-order logic. Because such independencies can be expressed in IF logic, a first-order IF language allows the formulation of a truth predicate for itself. Because Tarski's theorem is thus due to the expressive poverty of traditional logic (rather than its excessive deductive strength, as usually thought), definability problems are no obstacles to the explicit use of the notion of truth also in natural languages (cf. Hintikka, 1996; Sandu, 1998). Because much of the recent philosophical discussion of the notion of truth has in effect been prompted by Tarski's undefinability theorem, most of this discussion has to be reconsidered.

12. GTS, in the sense used here, is characterized by the definition of truth of S as the existence of a winning strategy for V in a semantical game G(S) associated with S. This meaning of GTS has to be distinguished from other uses of games and game theory in logic and linguistics, such as dialogical

games (including questioning games), games of formal proof, games of communication, and so on. There are interesting connections between these different kinds of games. For instance, games of formal proof can be considered as mapping the different courses that a play of a semantical game can take. Again, the strategies of deductive games are closely related to the strategies of the corresponding questioning games when all answers are known *a priori* to be true (i.e., games of pure discovery).

13. Historically, GTS was inspired by Ludwig Wittgenstein's notion of language game. Philosophically, these two share the idea that language-world relations are constituted by certain rule-governed human activities. This implies that intentional relations are dispensable in semantics. In neither kind of game are the moves normally made by speech acts or other language acts. Both games are objective in that their theory depends only on their rules.

In sum, GTS is not merely one branch of formal semantics among others. It is an approach to all semantics, based on the possibility of considering the language-world links as being mediated by games in the abstract sense of the mathematical theory of games.

See also: Boole and Algebraic Semantics; Compositionality: Philosophical Aspects; Donkey Sentences; Negation: Philosophical Aspects; Negation: Semantic Aspects; Truth Conditional Semantics and Meaning; Truth: Theories of in Philosophy.

Bibliography

- Cameron P & Hodges W (2001). 'Some combinatorics of imperfect information.' *Journal of Symbolic Logic* 66, 673–684.
- Hintikka J (1987). 'Game-theoretical semantics as a synthesis of truth-conditional and verificationist theories of meaning.' In Lepore E (ed.) *New directions in semantics*. London: Academic Press. 235–258.
- Hintikka J (1996). *The principles of mathematics revisited*. Cambridge, UK: Cambridge University Press.
- Hintikka J (1997). 'No scope for scope.' *Linguistics and Philosophy* 20, 515–544.
- Hintikka J (2002a). 'Negation in logic and natural language.' *Linguistics and Philosophy* 25, 585–600.
- Hintikka J (2002b). 'Quantum logic as a fragment of independence-friendly logic.' *Journal of Philosophical Logic* 31, 197–209.
- Hintikka J & Halonen I (1995). 'Semantics and pragmatics for why-questions.' *Journal of Philosophy* 92, 636–657.
- Hintikka J & Kulas J (1983). *The game of language*. Dordrecht: D. Reidel.
- Hintikka J & Kulas J (1985). *Anaphora and definite descriptions: Two applications of game-theoretical semantics*. Dordrecht: D. Reidel.

- Hintikka J & Sandu G (1989). 'Informational independence as a semantical phenomenon.' In Fenstad J E, Frolov I T & Hilpinen R (eds.) *Logic, methodology and philosophy of science VIII*. Amsterdam: North-Holland. 571–589.
- Hintikka J & Sandu G (1991). *On the methodology of linguistics: A case study*. Oxford: Basil Blackwell.
- Hintikka J & Sandu G (1996). 'Game-theoretical semantics.' In van Benthem J & ter Meulen A (eds.) *Handbook of logic and language*. Amsterdam: Elsevier. 361–410.
- Hodges W (1997). 'Compositional semantics for a language of imperfect information.' *Logic Journal of the IGPL* 5, 539–563.
- Sandu G (1998). 'IF logic and truth-definition.' *Journal of Philosophical Logic* 27, 143–164.
- Sandu G & Hintikka J (2001). 'Aspects of compositionality.' *Journal of Logic, Language, and Information* 10, 49–61.

Generative Grammar

J Rooryck, Leiden University, Leiden,
The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Like many linguistic terms, the phrase 'generative grammar' means different things to different practitioners of linguistics. The term 'grammar' itself can be loosely defined as the set of rules that accurately describe the combination of elements in a language. A properly 'generative' grammar is, roughly speaking, a grammar whose rules generate (i.e., produce) all and only the correct combinations of elements in a language. This definition corresponds to the use of generative grammar as a common noun. However, the term 'generative grammar' gradually received a much broader meaning as a proper noun, referring to the specific research program that is associated with the mentalist approach to language launched and developed by Noam Chomsky. In this article, the basic characteristics of both meanings of generative grammar are sketched, and the relationship between them is elucidated.

The narrower, common noun meaning of the term 'generative grammar' originates in mathematical recursive function theory and is also used in computer science. A generative grammar, then, is one of two types of formal grammars, generative and analytic. Such a formal grammar comprises a set of rules, which in turn define a possibly infinite set of finite strings composed of a set of fixed elements. These elements need not be words of natural language; they may be any kind of symbol.

A classical generative grammar contains four components: (1) a finite set of nonterminal symbols, (2) a start symbol contained in the set of nonterminal symbols, (3) a finite set of terminal symbols, and (4) a finite set of production rules that rewrite a (string containing at least one) nonterminal symbol into a string of terminal and/or nonterminal symbols. Rules

apply in any order and any number of times until all nonterminal symbols are rewritten as strings of terminal symbols. The set of strings with terminal symbols that can be generated this way is called the 'language' generated by the grammar. Generative grammars can be classified into four types conforming to the so-called Chomsky hierarchy. These types differ by an increasingly stricter format for production rules and concomitantly fewer formal languages. In this formal sense, the term 'grammar' is therefore used only by analogy with the term 'grammar' in natural languages.

Take a generative grammar that consists of the set of nonterminal symbols $\{X, Y\}$ with X the start symbol, the set of terminal symbols $\{a, b\}$, and the rules $X \rightarrow aYb$, $Y \rightarrow Xb$, and $Y \rightarrow ba$. The arrow inside the rules carries the instruction 'is rewritten as.' Applying $X \rightarrow aYb$, followed by an application of $Y \rightarrow Xb$ yields the intermediate string $aXbb$. This string still contains a nonterminal symbol. Therefore, it requires reapplication of the rule $X \rightarrow aYb$ (yielding $aaYbbb$) and subsequently the rule $Y \rightarrow ba$ to yield a string that consists solely of terminal symbols, in this case $aababbb$. This grammar also generates other strings (infinitely many, in fact), which together constitute the language of this particular generative grammar.

This formal apparatus provides a powerful tool for analyzing natural language. The set of terminal symbols, then, contains the vocabulary of a natural language, say, English. The set of nonterminal symbols contains word class labels (noun, verb, etc.), as well as larger syntactic units, such as noun phrases and sentences. The largest syntactic unit, the sentence, can be taken as the nonterminal start symbol. The production rules of a descriptively adequate grammar generate the set of sentences that an English speaker would understand as uniquely English, and they fail to generate sentences and phrases that are not judged to be proper English by such a speaker. An English

speaker will readily accept *the red book* as a grammatical unit, but not **the book red*. A more traditional description would then state that in English, adjectives end up in front of nouns and do not appear after them. In a generative grammar of English, this descriptive fact is a result of the order specified in the rewriting rule. In this rule, a noun phrase (NP) is rewritten as a string in which the Determiner (Det) is followed by the adjectival phrase (AP) and the noun (N) in that order: NP → Det AP N. Replacing the nonterminal symbols Det, AP, and N by *the*, *red*, and *book* respectively yields *the red book*. By contrast, the rewriting rule that makes adjectives end up after nouns (NP → Det N AP) is not part of the generative grammar of English. The grammar thus provides a fully explicit syntax of English, rather than the informal or implicit characterization often found in traditional grammars. English speakers have rules that allow them to generate an infinite number of new sentences by using a finite vocabulary, and to interpret entirely new combinations of words. A generative grammar can be recursive: the output of one application of a rewrite rule can serve as the input for a later application of the same rule. This mechanism of recursion is also active in natural language; it is responsible for such sequences as *the sister of the brother of the grandfather of the niece of an acquaintance of my aunt*, which can in principle be infinitely extended by recursion, limited only by performance factors.

Applying this line of thinking in a mentalist context, Noam Chomsky in the late 1950s revolutionized the way of looking at natural language. The study of language could reveal “abstract principles that govern its structure and use, principles that are universal by biological necessity and not mere historical accident, that derive from the mental characteristics of the species” (Chomsky, 1975: 3). The computationally ‘generative’ properties of the grammars of natural languages, and more in particular the property of recursion, is one instance of such abstract, universal, biologically necessary characteristics of the grammars developed by the human species. It should be noted, however, that it is perfectly possible to study properties of formal generative grammars outside of the mentalist context, i.e., generative grammars can be part of nonmentalist approaches to grammar.

Nevertheless, the school of thought on the mentalist nature of language developed by Chomsky and his followers came to be known – rather misleadingly and somewhat erroneously – under the popular name ‘generative grammar’ or ‘generative linguistics,’ this time used as a proper noun. In this much broader meaning, the name should be taken as no more than an expedient, historically motivated *pars pro toto*,

encompassing not only syntax but a mentalist model of grammar also comprising phonology, morphology, and semantics. Accordingly, this school of thought is not limited to finding the generative grammar generating all and only the grammatical sentences of a language in the formal sense defined earlier in this article. Perhaps even more surprisingly, however, the model of generative grammar (as a proper noun) developed in the Minimalist program does not even contain a generative grammar in the common noun sense. Rewriting rules have been entirely dispensed with in favor of more elementary, structure-building operations. Thus, the proper noun meaning of ‘generative grammar’ has entirely outgrown the common noun meaning. Understandably, then, Chomsky himself opposes the use of ‘generative grammar’ to describe his approach to language, preferring terms such as the rather laborious ‘principles and parameters framework’ or the succinctly elegant ‘biolinguistics.’ Despite such opposition, the term is in widespread use as representing the scientific paradigm that has dominated theoretical linguistics for the past 50 years. The remainder of this article is therefore devoted to characterizing generative grammar as a proper noun in the sense already defined.

The approach to language initiated by Chomsky is radically *internalist*. Language is viewed as a species-specific property, part of the mind/brain. The object of study of linguistics as a part of cognitive science is this mental faculty for language. Chomsky radically equates linguistic theory with the study of *I(nternal)-language*, language as an internal cognitive structure. A theory of grammar is a model of the speaker’s linguistic competence, part of the internal structure of his mind. A speaker of a language has internalized a system of rules relating sound to meaning in a particular way. The task of the theoretical linguist is to describe and explain the linguistic competence of a native speaker, the linguistic knowledge of language present in the minds of individuals. A linguist construes hypotheses concerning this internalized system. *E(xternal)-language* is not part of the theory of language and comprises sundry approaches to language as it is used in society, literature, or communication.

The fact that any child can acquire any of the 6000 or so languages in the world in a relatively short period of time and under imperfect input conditions gives rise to the *Immateness Hypothesis*. This hypothesis encapsulates the claim that some set of fundamental characteristics of all human languages must be the same, and part of the *Language Faculty*. The faculty of language is part of the human biological endowment. Children are born with a faculty of language in its Initial State, i.e., the state of

the faculty of language before it is exposed to the input of a particular language. This innate knowledge is sometimes called *Universal Grammar*. Language acquisition requires interaction with the environment, in the same way birds need interaction with the environment to 'learn' how to fly or sing. It is assumed that particular properties of Universal Grammar are parametrized, hence the term Principles and Parameters framework. Particular choices for each parameter of UG lead to specific grammars of individual languages. The Initial State is therefore progressively replaced during acquisition with grammars that come to resemble the Steady State, i.e., the internalized grammar of the language of a given adult speaker.

See also: Principles and Parameters Framework of Generative Grammar.

Bibliography

- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton.
 Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
 Chomsky N (1975). *Reflections on language*. New York: Pantheon.
 Chomsky N (1986). *Knowledge of language*. New York: Praeger.
 Chomsky N (1988). *Language and problems of knowledge: the Managua lectures*. Cambridge MA: MIT Press.
 Chomsky N (1995). *The Minimalist program*. Cambridge, MA: MIT Press.
 Chomsky N (2002). *On nature and language*. Cambridge: Cambridge University Press.
 Smith N V (1999). *Chomsky: ideas and ideals*. Cambridge: Cambridge University Press.

Generative Semantics

J D McCawley and R A Harris

© 2006 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition article by James D McCawley, volume 3, pp. 1398–1403, (c) 1994, Elsevier Ltd., with a foreword by Randy Harris.

Foreword (by Randy Harris)

There is little that can or should be added to the definitive epitome of generative semantics you are about to read, by James D. McCawley (1938–1999), except (1) a few words about the importance of McCawley to the movement, which is perhaps, less prominent in an article of his own authorship than it may have been from anyone else's pen, and (2) a few additional citations.

Each of the four main figures McCawley associates with generative semantics – George Lakoff (b. 1941), John Robert (Háj) Ross (b. 1938), Paul Postal (b. 1936), and himself – contributed very substantial elements to its identity, but McCawley embodied the approach, from his feet to his very lively eyebrows, and especially above. He was, in all senses of the phrase, its presiding genius. He helped bring it to life in lengthy, rollicking, mid-1960s telephone calls with Lakoff between Cambridge, Massachusetts, and Chicago. He supplied what many regarded as its strongest arguments and its most compelling analyses, some of which brought Postal into the program. He spent his entire career in the movement's epicenter, the University of Chicago.

He continued to publish comprehensive works in the generative semantics spirit long after the label had fallen into disrepute, especially *Syntactic Phenomena* (1993a) and *Everything That Linguists Have Always Wanted to Know about Logic* (1993b). He believed in generative semantics to the very end, not in all of its specific proposals (relentlessly honest, he cheerfully and publicly dropped analyses that no longer fit his evolving views and cheerfully welcomed views that did, no matter what their origin), and certainly not in the label itself (indeed, he renounced all theoretical labels), but in its substance.

Further reading in generative semantics include Lakoff (1971), McCawley (1976, 1979), Postal (1972), Ross (1972, 1973), Newmeyer (1980, McCawley cites the 1986 second edition; the 1980 first edition has more on generative semantics), Lakoff (1989), Harris (1993a, 1993b), Huck and Goldsmith (1996), and McCawley (1981). Also of note are the two festschrifts for McCawley: Brentari *et al.* (1992) and Zwicky *et al.* (1970/1992).

Generative Semantics (by James D McCawley)

The term 'generative semantics' (GS) is an informal designation for the school of syntactic and semantic research that was prominent from the late 1960s through the mid-1970s and whose best-known practitioners were George Lakoff, James D. McCawley, Paul M. Postal, and John Robert Ross.

GS Positions on Controversial Issues

The name GS gives undue prominence to one of many issues on which GS-ists took positions that conflicted with those of more orthodox generative grammarians, an issue that in hindsight seems arcane because it is intelligible only against the background of the once widely accepted assumption (shared then by GS-ists and their adversaries) that there must be a single level of linguistic structure for which it is appropriate to give a system of 'generative rules' (i.e., rules giving a complete specification of what structures are well-formed on that level) and to which all other levels of structure are related by 'interpretive rules.' The issue commemorated in the name GS was that of whether the privileged level was semantic structure (the GS position) or was a level of syntactic structure as distinct from semantic structure (the position of Chomsky and other 'interpretive semanticists'). The prominence that has been given to that arcane issue should not obscure the fact that GS-ists disagreed with other generative grammarians on many far more substantial issues, such as the following:

- a. Whether sentences were 'grammatical' or 'ungrammatical' in themselves rather than relative to (linguistic and extralinguistic) contexts and to possible interpretations. GS-ists rejected the then popular idea that a language can be identified with a set of sentences and took syntactic derivations as implying that the surface form in question was grammatical not absolutely but only relative to the meaning represented in its deep structure and to any contextual factors to which steps in the derivation are sensitive.
- b. The nature of semantic structure. GS-ists held that semantic structures have the same formal nature as syntactic structures, except for having semantic rather than morphological entities as their ultimate constituents, while interpretive semanticists either were reluctant to make any concrete claims about the nature of semantic structure (e.g., Chomsky, 1972: 137) or adopted a conception of semantic structure that differed considerably in formal nature from syntactic structure (e.g., Jackendoff, 1972).
- c. The nature of syntactic categories. Much work in GS attempted to reduce syntactic category distinctions to distinctions of logical category, supplemented by lexical 'exception features' (e.g., verbs and adjectives would both belong to the category 'predicate,' usually confusingly called 'V' by GS-ists, with adjectives differing from verbs in bearing a feature licensing the application of a transformation that inserts a copula), while other generative

grammarians took syntactic categories to have at most a tangential relation to semantic categories.

- d. The linguistic level or levels relevant to the choice of the lexical material of a sentence. One who holds that there is no level of syntactic deep structure as distinct from semantic structure is forced to recognize syntactic structures whose ultimate units are semantic rather than morphological in nature, such as a syntactic structure [Brutus DO SOMETHING_x (X CAUSE (BECOME (NOT (Caesar ALIVE)))))] underlying *Brutus killed Caesar*. (Here and below, capitalization is used as an informal way of representing semantic units corresponding roughly to the words in question.) GS-ists accordingly proposed transformations that combined semantic units into complexes that could potentially underlie lexical items, e.g., 'predicate-raising' (proposed in McCawley, 1968) adjoined a predicate to the immediately superordinate predicate, thus allowing the derivation of such complexes as NOT-ALIVE, BECOME-NOT-ALIVE (= *die*), BECOME-NOT (= *cease*), and CAUSE-BECOME-NOT-ALIVE or CAUSE-*die* (= *kill*). Intermediate derivational stages involving both lexical and semantic units (such as CAUSE-*die*) needed to be recognized in order to account for, e.g., the parallelism between idiomatic combinations with *come* (*come about*, *around*, *to...*) and their counterparts with *bring*: as Binnick (1971) noted, *bring* corresponded not to CAUSE plus some determinate complex of semantic material but to CAUSE plus *come*, irrespective of whether *come* was used as an independent lexical unit or as part of such combinations as *come about*. Consequently, lexical insertion could not be restricted to a single linguistic level: applications of certain transformations had to be interspersed between lexical insertions. The combinations that could be derived through the application of predicate-raising and other 'prelexical' transformation were supposed to underlie 'possible lexical items.' Since there are infinitely many such combinations but only finitely many actual lexical items in any given language, most correspond to no actual lexical item of the language and were supposed to reflect accidental gaps in the lexicon of the language.

Lexical decomposition analyses were criticized in such works as Fodor (1970), where it was argued that the simple and complex surface forms that supposedly corresponded to the same deep structure (e.g., *Brutus killed Caesar* and *Brutus caused Caesar to die*) did not in fact have the same meanings. It was noted subsequently (McCawley, 1978) that such

discrepancies in interpretation can be explained by a version of Grice's (1967/1989) maxim of manner according to which a simple surface form is preferred to a more complex alternative except when the referent is a peripheral instance of the category defined by the given semantic structure, e.g., using indirect means to cause someone to die would be a peripheral instance of the category defined by 'cause to cease to be alive' and thus would not be in the part of that category where *kill* would preempt the use of *cause to die*. (Syntactic analyses involving lexical decomposition also figured prominently in Gruber, 1965, a work that early GS-ists found congenial despite some important differences between its framework and theirs.)

GS Policies on the Conduct of Research

Of equal importance to these points of theory in their influence on the directions that GS research took and the reception that GS work received were several policies about the conduct of linguistic research, of which the following deserve mention here:

- a. A lack of concern about the compartmentalization of the parts of a linguistic analysis or of a linguistic theory, as contrasted with the concern among Chomskyan generative grammarians with the drawing of boundaries among, e.g., syntax, semantics, and pragmatics. One important facet of this lack of concern was an egalitarian position regarding the different kinds of data that had a bearing on a linguistic analysis: whereas most generative grammarians held that syntactic analyses needed to be supported by arguments in which only syntactic facts figured, GS-ists held that facts about truth conditions, possible denotations, etc., were as relevant as any other kind of facts to evaluating analyses that purported to specify how meanings corresponded to surface forms in the given language, and that supposed syntactic facts were usually at least partly semantic in nature, in that what a speaker of a language judges acceptable is not a sentence in itself but that sentence relative to an assumed understanding of it. Another facet of this policy was GS-ists' insistence that all parts of a linguistic analysis were subject to the same standards of explicitness, simplicity, and factual accuracy, irrespective of how one might wish to demarcate syntax and semantics; by contrast, interpretive semantics has come only gradually and often grudgingly to subject the semantic parts of analyses to the same standards of appraisal as the syntactic parts.
- b. Rejection of the dogma of generative grammar that a fixed notational system is essential to a linguistic theory and that precision can be achieved only by formulating one's analyses in the privileged notational system.
- c. Adoption of a 'static' conception of linguistic rules: rules were thought of not in terms of the popular metaphor of assembling linguistic structures and converting structures on one level into corresponding structures on another, but as 'derivational constraints,' that is, as specifications of what a structure may or may not contain and of how a structure on one level may not contain and of how a structure on one level may or must differ from the corresponding structure on another level. This difference in the conception of rules resulted in difference with regard to what theoretical notions posed 'conceptual problems' (Laudan, 1976) for each approach; thus GS-ists readily accepted rules that specified relations among nonadjacent levels of structure (what Lakoff, 1970b dubbed 'global rules'), a notion that was unproblematic from their conceptual vantage point but outlandish from the vantage point of the 'operation' metaphor for rules, while rejecting the idea of ordering among rules, a notion that was unproblematic for those who accepted the 'operation' metaphor but was difficult to make coherent with the GS conception of rules as derivational constraints.
- d. Disdain for those concerns of Chomskyan generative grammarians that had little connection with linguistic facts or with detailed linguistic description, such as mathematical models and speculation about the extent to which linguistic structure is biologically determined. While GS-ists were receptive to the idea that linguistic structure is profoundly influenced by neuroanatomy, they demanded (e.g., Lakoff, 1974: 171) that claims to that effect be backed up with solid linguistic and solid biology rather than with what they dismissed as arguments from ignorance (i.e., hasty leaps from one's failure to see how some characteristic of languages could be learned to the conclusion that it must be innate).
- e. Eagerness to put in practice in their professional lives many of the ideas of the 1960s counterculture, such as policies of antiauthoritarianism, antielitism, and demystification of science and scholarship, and a belief that one's work should be pleasurable. One of many facets of the GS ethos that these policies helped to shape is what Newmeyer (1986: 133) has disparaged as 'data-fetishism': joy in the unearthing of novel and intriguing facts for which one is not yet in a position to provide a satisfactory analysis; GS-ists, by contrast,

regarded Chomskyan generative grammarians as 'scientific Calvinists' (McCawley, 1980: 918).

Prominent and Influential Analyses Proposed within the GS Approach

Kuhn (1970) notes that one major component of the paradigm of a scientific community is a set of 'exemplars': prestigious solutions to problems, presented to neophytes in the field as paragons of good science, and serving as models for solutions to new problems. (For discussion of the history of generative grammarians' analyses of English auxiliary verbs in terms of Kuhnian notions such as 'paradigm' and 'exemplar,' see McCawley, 1985.) The exemplars for the GS community included a number of analyses that bore an intimate relation to central tenets of GS, for example, lexical decomposition analyses such as were discussed in, First section of this article, and analyses of quantified expressions as being external to their host sentences in deep structure (e.g., *John has read many books* = *many books* + *John has read x*) and as being moved into their surface positions by a transformation of Quantifier-Lowering (QL). (The term QL was in fact applied indiscriminately to a variety of transformations that differed according to the specific deep structure that was assumed; proposals differed with regard to whether just a quantifier or a whole quantified expression was external to the host S, what filled the deep structure position into which the quantified expression was to be moved, and where the quantifier or quantified expression was in relation to the host S.) The best-known arguments given for a QL analysis consisted in demonstrations that the many syntactic rules that were problematic when applied to structures that contained quantified elements became unproblematic if the deep structure position of a quantified expression was external to its host sentence and consequently (in virtue of the principle of the cycle) the rule had as its domain of application a structure that does not contain the quantified expression; for example, this view of the interaction between QL and the transformation of 'Reflexivization' explained why such pairs of sentences as *Every philosopher admires himself* and *Every philosopher admires every philosopher* differed in meaning in the way in which they did, and why reflexivization was applicable only in the derivation of the former. A thorough survey of arguments for a QL analysis is given in McCawley (1988: Ch. 18).

Several other GS exemplars were in fact as consistent with the substantive claims of interpretivist transformational grammar as with those of GS, but were embraced by GS-ists and rejected by interpretive semanticists as much because of policies on

the conduct of research (see previous section) as because of any points of linguistic theory, or simply because of the historical quirk that a particular idea occurred to a member of the one camp before it occurred to any of his counterparts in the other camp. One such exemplar is the analysis of English auxiliary verbs as being verbs that take nonfinite sentential complements in the manner of such verbs as *seem* (Ross, 1969; McCawley, 1971), which Pullum and Wilson (1977) subsequently argued for from within an interpretive semantic framework. (A similar treatment of auxiliary verbs is found in Jespersen, 1937: 92). A second was the proposal (McCawley, 1970, subsequently disavowed by the author) that English had underlying verb-subject-object (SVO) word order, a hypothesis that is, if anything, harder to reconcile with the assumptions of GS than with those of interpretive semantics in view of the dubious nature of the assumption that the order of elements is significant in semantic structure; by contrast, there is no general policy in interpretive semantic versions of generative grammar against discrepancies between deep and surface constituent order, and indeed languages with surface VSO word order are commonly analyzed by interpretive semanticists as having deep VSO word order. Another such exemplar is the 'performative analysis' (Ross, 1970), in which sentences are assigned underlying structures in which a 'hyper-sentence' (Sadock, 1969, 1974) specifies the illocutionary force of the sentence, e.g., *Birds fly* would have an underlying structure of the form [*I tell you [birds fly]*].

The History of GS

The term 'generative semantics' first appears in Lakoff (1963/1976), a work that antedates the development of the Katz-Postal-*Aspects* approach to syntax and prefigures some of Lakoff's subsequent GS work. GS originated in attempts by Postal and Lakoff to exploit novel possibilities that were opened up by the revisions of the transformational syntactic framework proposed in Katz and Postal (1964) and Chomsky (1965) and to fill gaps in the evolving framework. For example, Lakoff's Ph.D. thesis (Lakoff, 1965/1970) originated as an attempt to develop a precise and coherent account of the way in which a lexical item could affect the applicability of transformations to structures containing the given item, and thereby to put on a solid footing those analyses in Chomsky (1965) in which the choice of lexical items affected the possibilities for derivations. In the course of providing such a theory of 'rule model theory.) Coincidentally, the radical revisions that interpretive semanticists were making in their

versions of generative syntactic theory included the adoption of the 'X-bar' conception of syntactic categories, which identified two of the factors that affect the syntactic behavior of a linguistic unit, namely, the difference between a word unit and a phrasal unit, and the part of speech of the unit or of its head. Once a descriptive framework was available that allowed linguistic generalizations to be stated in terms of those factors, considerable progress was made in the analysis of the many syntactic phenomena in which those factors play a role.

No important tenets of GS rule out the adoption of a conception of syntactic categories as defined by these factors *in addition to* logical categories, and indeed a conception of syntactic categories as reducible to those and other factors (with logical category being merely one of several factors that influence a unit's syntactic behavior) is adopted in McCawley, 1977/1982 and subsequent works. However, in the 1960s and early 1970s, an assumption shared by GS-ists and interpretive semanticists impeded GS-ists from adopting such a conception of categories, namely the assumption that syntactic categories must remain constant throughout derivations: a word (with a determinate part of speech) that replaced a complex of semantic material (thus, a unit not having a part of speech) could not differ in category from the replaced unit and thus parts of speech could not be part of the category system. (The widespread misconception that GS analyses allowed linguistic units to change category in the course of derivations in their analysis of, for example, nominalizations overlooks the fact that, according to GS-ists' assumptions, verbs and their nominalizations belonged to the same category. Anyway, analyses of any kind in which the verb *invent* is a constituent of the noun *invention* are not committed to any assumption that the former changes its category in the derivation of the latter: it is whatever it is, regardless of what it is contained in.) Since interpretive semanticists did not require that deep structures match semantic structures (and indeed took delight in arguing that they did not match), there was no obstacle to their having the same categories in deep as in surface structures while drawing the full range of category distinctions provided for by X-bar syntax. The interpretive semantic research program was thus able to become 'progressive' in the sense of Lakatos (1978) because of something extraneous to the issues that were the loci of the substantive disputes between GS and interpretive semantics.

Bibliography

Bach E & Harms R T (eds.) (1968). *Universals in linguistic theory*. New York: Holt, Rinehart and Winston.

- Binnick R I (1971). 'Come' and 'bring.' *Lin* 2, 260–267.
- Brentari D, Larson G N & MacLeod L A (eds.) (1992). *The joy of grammar: a festschrift in honor of James D. McCawley*. Philadelphia: Benjamins.
- Chomsky N A (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky N A (1972). *Studies on semantics in generative grammar*. The Hague: Mouton.
- Fodor J A (1970). 'Three reasons for not deriving 'kill' from 'cause to die'.' *Lin* 1, 429–438.
- Grice H P (1967). 'Logic and conversation.' In Grice H P (ed.) (1989) *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Gruber J S (1965). 'Studies in lexical relations.' In Gruber J S (ed.) (1976) *Lexical structures in syntax and semantics*. North-Holland: Amsterdam.
- Harris R (1990). *The generative heresy*. Troy, NY: Rensselaer Polytechnic Institute, Unpublished Ph.D. diss.
- Harris R A (1993a). 'Generative semantics: secret handshakes, anarchy notes, and the implosion of ethos.' *Rhetoric Review* 23, 125–160.
- Harris R A (1993b). *The linguistics wars*. New York: Oxford University Press.
- Huck G J & Goldsmith J A (1996). *Ideology and linguistic theory: Noam Chomsky and the deep structure debates*. London: Routledge.
- Jackendoff R S (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Jespersen O (1937). *Analytic syntax*. London: Allen and Unwin.
- Katz J J & Postal P M (1964). *An integrated theory of linguistic description*. Cambridge, MA: MIT Press.
- Kuhn T (1970). 'Postscript.' *The structure of a scientific revolution* (2nd edn.). Chicago, IL: University of Chicago Press.
- Lakatos I (1978). 'Falsification and the methodology of research programmes.' In Lakatos I & Musgrave A (eds.) *criticism and the growth of knowledge*. Cambridge: Cambridge University Press.
- Lakoff G (1963). 'Toward generative semantics.' In McCawley J (ed.) (1976) *Notes from the linguistic underground (syntax and semantics 7)*. New York: Academic Press.
- Lakoff G (1970a). *Irregularity in syntax*. New York: Holt, Rinehart and Winston.
- Lakoff G (1970b). 'Global rules.' *Lg* 46, 627–639.
- Lakoff G (1971). 'On generative semantics.' In Steinberg D D & Jakobovits L A (eds.) *Semantics: an interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge, UK: Cambridge University Press. 232–296.
- Lakoff G (1974). 'Interview conducted by H Parret.' In Parret H (ed.) *Discussion language*. The Hague: Mouton.
- Lakoff G (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.
- Lakoff R T (1989). 'The way we were; or, the real actual truth about generative semantics: a memoir.' *Journal of Pragmatics* 13, 939–988.
- Laudan L (1976). *Progress and its problems*. Berkeley, CA: University of California Press.

- Levi J N (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.
- McCawley J D (1968a). 'The role of semantics in a grammar.' In Bach E & Harms R T (eds.).
- McCawley J D (1968b). 'Lexical insertion in a transformational grammar without deep structure.' *Chicago Linguistic Society Papers* 4, 71–80.
- McCawley J D (1971). 'Tense and time reference in English.' In Fillmore C & Langendoen D T (eds.) *Studies in linguistic semantics*. New York: Holt, Rinehart and Winston.
- McCawley J D (1975). 'Review of Chomsky 1972.' *Studies in English Linguistics* 5, 209–311.
- McCawley J D (1976). *Grammar and meaning*. New York: Academic Press.
- McCawley J D (1977). 'The nonexistence of syntactic categories.' In *Second Annual Metatheory Conference Proceedings*. East Lansing, MI: Michigan State University.
- McCawley J D (1978). 'Conversational implicature and the lexicon.' In Cole P (ed.) *Pragmatics* (Syntax and Semantics 9). New York: Academic Press.
- McCawley J D (1979). *Adverbs, vowels, and other objects of wonder*. Chicago: University of Chicago Press.
- McCawley J D (1980). 'Review of 1st edn. of Newmeyer 1986.' *Linguistics* 18, 911–930.
- McCawley J D (1981). 'Review of F. Newmeyer, *Linguistic theory in America*.' *Linguistics* 18, 911–930.
- McCawley J D (1982). *Thirty million theories of grammar*. London: Croom Helm.
- McCawley J D (1985). 'Kuhnian paradigms as systems of markedness conventions.' In Makkai A & Melby A (eds.) *Linguistics and philosophy: studies in honor of Rulon S. Wells*. Benjamins, Amsterdam.
- McCawley J D (1988). *The syntactic phenomena of English* (2 vols). Chicago, IL: University of Chicago Press.
- McCawley J D (1993a). *Everything that linguists have always wanted to know about logic (but were ashamed to ask)* (2nd edn.). Chicago: University of Chicago Press.
- McCawley J D (1993b). *The syntactic phenomena of English* (2nd edn., 2 vols). Chicago: University of Chicago Press.
- Musgrave A (1976). 'Why did oxygen supplant phlogiston? Research programmes in the Chemical Revolution.' In Howson C (ed.) *Method and appraisal in the physical sciences*. Cambridge: Cambridge University Press.
- Newmeyer F J (1980). *Linguistic theory in America*. New York: Academic Press.
- Newmeyer F J (1986). *Linguistic theory in America* (2nd edn.). Orlando, FL: Academic Press.
- Postal P M (1972). 'The best theory.' In Peters P S (ed.) *The goals of linguistic theory*. Englewood Cliffs, NJ: Prentice Hall.
- Postal P M (1974). *On raising*. Cambridge, MA: MIT Press.
- Pullum G K & Wilson D (1977). 'Autonomous syntax and the analysis of auxiliaries.' *Lg* 53, 741–788.
- Ross J R (1969). 'Auxiliaries as main verbs.' *Studies in Philosophical Linguistics* 1, 77–102.
- Ross J R (1970). 'On declarative sentences.' In Jacobs R & Rosenbaum P S (eds.) *Readings in English transformational grammar*. Waltham, MA: Ginn.
- Ross J R (1972). 'Doubl-ing.' In Kimball J (ed.) *Syntax and semantics*, vol. 1. New York: Seminar Press. 157–186.
- Ross J R (1973). 'Slifting.' In Gross M, Halle M & Schützenberger M P (eds.) *The formal analysis of natural languages*. The Hague, The Netherlands: Mouton. 65–121.
- Sadock J (1969). 'Hypersentences.' *Paper in Linguistics* 1, 283–371.
- Sadock J (1974). *Toward a linguistic theory of speech acts*. New York: Academic Press.
- Zwicky A M, Salus P H, Binnick R I & Vanek A L (eds.) (1992). *Studies out in left field: defamatory essays presented to James D. McCawley. Current inquiry into language and linguistics* (vol. 4). Amsterdam: Benjamins.

Generic Reference

G Carlson, University of Rochester, Rochester, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Forms of Generic Reference

Generic reference is the term commonly used to describe noun-phrase reference in sentences that express generalizations. Some common examples are found in (1)–(3):

- (1) *Potatoes* are native to South America.
- (2) *The lion* is a fearsome beast.
- (3) *A pencil* is used for writing.

Generic reference is usually understood as making reference to kinds of things (*see also Natural Kind Terms*). When we speak of 'kinds,' we intend a classification system that is based on the denotations of nominal expressions, or sortals, of the language (for one view, *see Gupta, 1980*).

It is now commonly accepted that reference is not only limited to individuals or pluralities of individuals, but also to kinds or types of things as well. This is most evident in noun phrases of the type "this kind of animal," which evidences an overt postdeterminer of the class of 'kind,' 'type,' 'sort,' 'species,' and so on.

- (4) *This kind of animal* hibernates in the winter.

These kind-referring phrases can appear in quantified contexts as well. The analysis then is that the quantifier ranges over kinds of things, just as it ranges over individuals in the more usual instances.

- (5) *Three kinds of swallows* are found in the northeastern United States.
- (6) *Every type of tree* exchanges carbon dioxide for oxygen.

When the postdeterminer element is removed, there remains the possibility of interpreting the noun phrase as referring to or quantifying over kinds. This normally results in a type/token ambiguity. For instance, in (7) one could be talking about individual flowers in a given context or kinds of flowers; similarly for (8). This reading is called a *taxonomic reading* in Krifka *et al.* (1995).

- (7) Sharon photographed *every flower*.
- (8) *Several penguins* inhabit this frozen wilderness.

Examples such as (7) and (8) are ambiguous between a 'kind' reading and the more common individual reading. On the taxonomic reading, the denotation of the head noun is partitioned into subkinds, though this is done contextually since there are multiple ways to naturally partition any domain. For instance, automobiles can be partitioned according to body style (sedan, sports car, station wagon, etc.) or by manufacturer (BMW, Mazda, Volvo, etc.), among other ways. It is commonly noted that if one takes a mass term and syntactically treats it as a count term, by pluralizing it or pairing it with a determiner that selects for singular count nouns only, a taxonomic reading may emerge. Thus, in (9) we are speaking of kinds of wine, and in (10) of kinds of metal:

- (9) *Three wines* are stored in the cellar.
- (10) *Every metal* conducts electricity to some degree.

Another means by which kinds are referred to in natural language is by definite singular noun phrases. In English, this has a stylistically technical tone, but this is not a general feature of other languages. Three possible examples are:

- (11) *The computer* has changed society in many ways.
- (12) Scientists have now documented the entire life cycle of *the three-toed sloth*.
- (13) *The self-cleaning plow* was invented by John Deere.

These exemplify the definite generic on the most natural readings of the noun phrases in these examples. This reading appears in addition to the much

more frequent individual-denoting use of the definite article, and often results in ambiguity. Generally unambiguous is the use of the definite article with only an adjective (e.g., "*The rich* are often oppressors of *the poor*"). Other types of definite kind reference include uses of the proximal and distal demonstratives (*this*, *that*) in the plural. The demonstrative is not, on one interpretation, an actual indexical; instead, it colloquially conveys an emotional attitude toward the kind (Bowdle and Ward, 1995). It appears to be the same use of the demonstrative as when it accompanies a proper name (e.g., "*That Roberto* has done it again").

- (14) *Those spotted owls* (i.e., the kind spotted owl) are constantly being talked about by environmentalists.
- (15) Who invented *these computers*, anyway?

In addition, there are noun phrases that employ adjectives like 'typical,' 'average,' or 'normal,' which have a kind-reference reading, as in "*Your typical businessperson* takes eight plane trips per year." Supplementing definite generics are the consciously introduced Latinate natural kind names, lacking a definite article, which always have an elevated scientific tone no matter the language. This includes examples like '*felis domesticus*' (cat) or '*acer saccharum*' (sugar maple tree). These are unambiguous and always denote kinds. Though not of consciously Latinate origin, the use of 'man' in English as a generic functions in much the same way.

Beyond these are additional means of kind reference in natural language. Bare plurals – that is, plural noun phrases lacking a determiner or quantifier element, at least on one reading – may refer to kinds. The following are three examples:

- (16) *Airplanes* have made intercontinental travel a common event.
- (17) *Lions* once ranged from the tip of Africa to eastern Siberia.
- (18) *Hunting dogs* are most closely related to wolves.

Functioning much the same as bare plurals are undetermined mass expressions (in French, the definite article must be employed), which allow for generic reference to unindividuated domains.

- (19) *Water* is a liquid. (cf. Fr. "*L'eau est un liquide.*")
- (20) *Hydrogen* is the most common element in the universe.

Finally, the singular indefinite article allows for a generic usage, as in the following:

- (21) *A triangle* has three sides.
- (22) *A potato* contains vitamin C.

The bare plural and the indefinite singular are commonly distinguished from the definite singular in English in that the former two usually allow for additional descriptive material in the form of modification, whereas the noun phrases in the definite generic instance are much more limited.

- (23) *A cake without baking powder/Cakes without baking powder/??The cake without baking powder fails to rise properly in the oven.*

Unlike the bare plurals or indefinite singulars, the definite singular is basically limited to expression of well-established kinds, those taken to be already familiar from one's background knowledge. Furthermore, as Vendler (1971) notes, it does not appear they can be 'too general.' Thus, alongside 'the parabola' and 'the circle,' one does not find generic reference to 'the curve.' Currently, a full account of these facts is lacking.

Cross-linguistically, generic reference is carried out by noun phrases with definite and indefinite articles and with determinerless expressions quite generally. In languages without articles, the determinerless form typically has a generic interpretation in at least some sentence positions. While in English the plural form of the definite has generic reference only marginally at best, in German, which is closely related to English, the plural definite may take generic reference quite easily (Delfitto, 1998). If there are languages with articles or other determiners specific to generic reference, they have yet to be brought to general attention, or they may not exist at all.

It is important to distinguish generic reference from the type of sentence in which the expression appears. While generic reference takes place most commonly within the context of a generic or habitual sentence, not all generic or habitual sentences have a noun phrase with generic reference, and generic reference may take place within sentences that are episodic or that make reference to specific events. The clearest examples of this are sentences with the definite singular generic exhibiting the *avant-garde* reading (Krifka *et al.*, 1995). Consider the following example:

- (24) *The horse arrived in the New World around 1500.*

This means that some horses were introduced about that time, but implies that the event was the first time any modern horses had been in that area. To observe a shipment of horses arriving in the western hemisphere in 1980 and use (24) modified by "in 1980" to describe that event would be inappropriate. Other instances where there is kind-reference in episodic sentences, on at least one reading, include the following three examples:

- (25) Today, Professor James lectured to us on the history of *dinosaurs*.
 (26) Marconi invented *the radio*.
 (27) *Monkeys* evolved from *lemurs*.

Theory of Generic Reference

While most semanticists agree that at least certain noun phrases refer to (or quantify over) kinds of things, there is a tradition in which apparent kind reference is treated in terms of quantification over individuals. Stebbings (1930), for instance, suggests that the sentence "The whale is a mammal" expresses a universal proposition (similar to "All whales are mammals"), as does "Frenchmen are Europeans." Russell comments that the sentence "Trespassers will be prosecuted" "means merely that, if one trespasses, he will be prosecuted" (1959: 102), which reduces the analysis of the apparent kind reference (*trespassers*) to an indefinite in a conditional. However, Moore (1944), in response to Russell's theory of descriptions, cites examples like "The triangle is a figure to which Euclid devoted a great deal of attention" and "The lion is the king of beasts," both of which convincingly resist implicit quantificational or conditional analyses.

The most convincing evidence for kind reference in the semantics stems from predicate positions that select for something other than individuals and pluralities of individuals and that readily accept the types of noun phrases reviewed earlier. These are called kind-level predicates. Examples (26) and (27) contain kind-level predicates. While an individual might invent something, the direct object must express a kind of thing and not a particular individual or set of individuals. The verb 'evolve' relates species and other levels of biological classes to other such classes, but not individuals to individuals. The following are other examples of kind-level predicates:

- (28) Dogs are *common/widespread/rare*.
 (29) Insects are *numerous*.
 (30) The elm is *a type/kind of tree*.
 (31) The gorilla is *indigenous to Africa*.
 (32) The Chevrolet Impala *comes in 19 different colors*.

Kind-level predicates are relatively infrequent in any language. Most predicates fall into the classes of either individual level or stage level. Roughly speaking, stage-level predicates speak of highly temporary events and states, such as running across a lawn, eating a sandwich, or being asleep. Individual-level predicates, on

the other hand, speak of more permanent states of affairs, such as knowing French, liking the opera, or being intelligent. Typically, the predicates of a habitual or generic sentence (“x cooks wonderfully”) are individual level. These are discussed in more detail in Carlson (1980), Kratzer (1995), Fernald (2000), and by others. Both stage-level and individual-level predicates select for noun phrases that denote individuals and pluralities of individuals. However, kind-denoting expressions appear with these predicates naturally as well. With both stage-level and individual-level predicates, a quantificational analysis of kind-denoting phrases (quantifying over individuals of that kind) becomes easily possible.

The kind-level predicates do not typically allow for the use of the indefinite singular. An example like (33) is generally deemed not very acceptable:

- (33) ?A lion is a species of animal. (cf. *the lion, lions*)

A continuing controversy centers on the analysis of the English bare plural construction, which has an unlimited distribution in comparison to bare plurals in other languages with articles, such as Spanish or Italian (e.g., Laca, 1990). English bare plurals appear to have different interpretations in different contexts. With individual-level predicates, they have a general interpretation, one that is quantificationally similar to ‘all’ or ‘most.’

- (34) Cats (roughly, all or most cats) sleep a lot.

- (35) Hammers (roughly, all or most hammers) are used for driving nails.

On the other hand, bare plurals also have an existential interpretation in other contexts that is similar to “some” in force.

- (36) *Priceless works of art* (i.e., some works) were delivered to the museum yesterday.

- (37) The rioters threw *stones* through *shop windows*, shattering them.

The primary question is whether in these instances, as well, the bare plural construction is kind-denoting, as most believe it is with kind-level predicates. Carlson (1980) and Chierchia (1998) argue that such a unified analysis is not only possible but also desirable, and both present analyses showing how it can be accomplished. However, others argue that more adequate insight can be gained through an analysis that differentiates true instances of kind reference from those instances where bare plurals appear with individual-level and stage-level predicates and that a quantification over individuals approach is better taken (see Wilkinson, 1991; Diesing, 1992; Krifka *et al.*, 1995).

See also: Natural Kind Terms.

Bibliography

- Bacon J (1971). ‘Do generic descriptions denote?’ *Mind* 82, 331–347.
- Bowdle B & Ward G (1995). ‘Generic demonstratives.’ In *Proceedings of the Twenty-First Annual Meeting of the Berkeley Linguistics Society*, 32–43.
- Burton-Roberts N (1976). ‘On the generic indefinite article.’ *Language* 52, 427–448.
- Carlson G (1980). *Reference to kinds in English*. New York: Garland.
- Carlson G & Pelletier F J (eds.) (1995). *The generic book*. Chicago: University of Chicago Press.
- Chierchia G (1995). ‘Individual level predicates as inherent generics.’ In Carlson G & Pelletier F J (eds.) *The generic book*. Chicago: University of Chicago Press. 176–223.
- Chierchia G (1998). ‘Reference to kinds across languages.’ *Natural Language Semantics* 6, 339–405.
- Cohen A (2001). ‘On the generic use of indefinite singulars.’ *Journal of Semantics* 18, 183–209.
- Dahl Ö (1975). ‘On generics.’ In Keenan E (ed.) *Formal semantics of natural language*. Cambridge: Cambridge University Press. 99–111.
- Delfitto D (1998). ‘Bare plurals.’ In Everaert M & van Riemsdijk H (eds.) *Encyclopedia of syntactic case studies*. Wassenaar: SynCom Group (electronic version 61pp.).
- Delfitto D (2002). *Genericity in language*. Alessandria: Edizioni dell’Orso.
- Diesing M (1992). *Indefinites*. Cambridge, MA: MIT Press.
- Fernald T (2000). *Predicates and arguments*. Oxford: Oxford University Press.
- Greenberg Y (2003). *Manifestations of genericity*. New York: Routledge.
- Gupta A (1980). *The logic of common nouns*. New Haven, CT: Yale University Press.
- Kratzer A (1995). ‘Individual level predicates vs. stage level predicates.’ In Carlson G & Pelletier F J (eds.) *The generic book*. Chicago: University of Chicago Press. 125–175.
- Krifka M, Pelletier F J, Carlson G, ter Meulen A, Chierchia G & Link G (1995). ‘Genericity: an introduction.’ In Carlson G & Pelletier F J (eds.) *The generic book*. Chicago: University of Chicago Press. 1–124.
- Laca B (1990). ‘Generic objects: some more pieces of the puzzle.’ *Lingua* 81, 25–46.
- Landman F (1991). *Structures for semantics*. Dordrecht: Kluwer.
- Moore G E (1944). ‘Russell’s “Theory of descriptions.”’ In Schilpp P (ed.) *The philosophy of Bertrand Russell*. New York: Tudor Publishing. 175–226.
- Ojeda A (1993). *Linguistic individuals*. Stanford, CA: CSLI.
- Russell B (1959). *The philosophy of logical atomism*. Minneapolis: University of Minnesota.
- Schubert L K & Pelletier F J (1989). ‘Generically speaking, or, using discourse representation theory to interpret generics.’ In Chierchia G, Partee B & Turner R (eds.)

- Property theory, type theory, and semantics 2: Semantic issues.* Dordrecht: Kluwer. 193–268.
- Stebbins S (1930). *A modern introduction to logic.* London: Methuen.
- Vendler Z (1971). ‘Singular terms.’ In Steinberg D & Jakobovits L (eds.) *Semantics: an interdisciplinary reader*

- in philosophy, linguistics, and psychology.* Cambridge: Cambridge University Press. 115–133.
- Wilkinson K (1991). ‘Studies in the semantics of generic NP’s.’ Ph.D. diss., University of Massachusetts, Amherst.

Grammatical Meaning

Ö Dahl, Stockholm University, Stockholm, Sweden

© 2006 Elsevier Ltd. All rights reserved.

Grammatical meaning is usually seen as opposed to lexical meaning. Grammatical meaning thus ought to include any aspect of linguistic meaning that is due to the grammatical structure of an expression rather than to the choice of lexical items. The variation in the definitions found in the literature suggest, however, that the notion is not wholly well understood. Consider, for illustration, the following sample formulations: ‘the part of meaning that varies from one inflectional form to another’; ‘the meaning of a word that depends on its role in a sentence’; ‘the meaning of an inflectional morpheme or of some other syntactic device, as word order.’

A suitable point of departure for a discussion of the notion of grammatical meaning is the classic treatment by Edward Sapir (1921, Chapter V). Although the term ‘grammatical meaning’ itself is not used there, the topic, denoted in the chapter heading as ‘grammatical concepts,’ is the same. According to Sapir’s initial taxonomy, concepts used in language are either ‘concrete’ or ‘relational.’ This coincides, more or less, with the lexical:grammatical distinction. The use of the terms ‘concrete’ and ‘relational’ is thus somewhat different from what is usual. Sapir actually never gives an explicit definition of the terms but comments on his example ‘The farmer killed the duckling’ as follows: “A rough and ready analysis discloses here the presence of three distinct and fundamental concepts” (p. 82) – these are expressed by the three lexical words *farmer*, *killed*, and *duckling* – “that are brought into connection with each other in a number of ways.”

Apparently, then, the relational concepts, which are in this particular case, definiteness of reference, singularity, declarative modality, and ‘subjectivity’ and ‘objectivity’ (meaning the roles as subject and object), and which are expressed through grammatical morphemes or word order, are responsible for the connections between the concrete elements.

However, it is really only the last two, which correspond to the syntactic functions of the two nouns in the sentence, that are relational in the sense of specifying relations between the lexical elements. It is more difficult to understand why, for instance, singularity, ‘expressed by lack of plural suffix -s . . . and by suffix -s in the following verb’ has to be regarded as relational. Moreover, as Sapir notes, the cardinality of referents would not be systematically indicated in many languages, and can be expressed also by lexical means, e.g., by a numeral. The necessary conclusions are that not all relational (i.e., grammatical) concepts are equally essential in language. Sapir is thus led to postulate a category of ‘concrete relational concepts,’ which can vary from language to language, not only in their expression, but also as to their concreteness. Relational concepts such as subjectivity and objectivity on the other hand are not subject to this variation: “The fundamental syntactic relations must be unambiguously expressed” (p. 94). It does seem that Sapir gets a bit entangled in his own reasoning here. On the one hand, he no is longer able to use the distinction between concrete and relational as an explanation for what is grammatical and what is lexical in languages; rather, he takes the grammatical status of a concept as an indicator of its concreteness or relationality, thus opening himself to allegations of circularity. On the other hand, the unequivocally relational concepts, which have to be expressed in languages, no longer display themselves as clearly semantic: he himself speaks of ‘fundamental syntactic relations.’ Sapir’s predicament does reflect the complexity of what Crystal (1997: 107) aptly calls “an area of study that falls uneasily between semantics and grammar.”

Much of the uneasiness is related to the fact that distinguishing the contributions of grammar and lexicon to meaning is no less problematic than separating the roles of genes and environment in an individual. A neat division between grammatical and lexical meaning presupposes that linguistic expressions are like buildings where the building blocks (the lexical items) are independent of the cement

(the grammar). But there is always a restricted number of ways in which a lexical item can be combined with others, and the meaning of the lexical item can often be expressed only in terms of what the resulting combination means. For instance, it is rather difficult to explain the meaning of a verb such as 'borrow' without placing it in a construction such as 'NP borrows NP from NP,' or to explain the meaning of 'father' without mentioning the persons involved in the 'x is the father of y' relation. In formal semantic frameworks such as Montague Grammar, the meanings of relational terms such as 'borrow' and 'father' are specified in terms of functions (in the mathematical sense of 'mapping') from the meanings of the argument expressions to the meaning of the whole. By abstracting away from the content of the specific lexical items, we may arrive at more general constructions, such as 'NP Verb NP Prep NP,' which, in formal semantic terms, have to be interpreted as second-order functions (that is, functions that take other functions as their arguments). It is at the level of constructions that the interface between grammar and meaning has to be sought, as has been argued by proponents of Construction Grammar, and the grammatical meaning of linguistic elements in general can only be understood in terms of their roles in larger constructions.

Grammatical markers and features such as word order and prosody undoubtedly serve the function of making it easier to grasp the structure of a complex expression, in particular the hierarchical relationships between elements. However, the choice

of grammatical elements such as inflectional morphemes may also depend on factors that have little to do with the structure of a sentence and sometimes directly reflect extralinguistic features of the situation or of referents of the discourse. This makes it difficult to uphold a thesis that the meaning of inflectional markers (and of grammatical morphemes in general) is different in kind from that of lexical morphemes.

Rather, inflectional markers differ from lexical morphemes in the role their meanings play in the speech act. A past tense marker in a language such as English normally does not have the function of informing the listener that the event or state referred to took place in the past; rather, it is an obligatory feature (except in some specific styles or genres) of any verb that refers to a state-of-affairs in the past. A related property of inflectional markers is that it is typically difficult or impossible to focus on them or, what can be seen as a special case of focusing, to negate them separately.

See also: Syntax-Semantics Interface.

Bibliography

- Contini-Morava E & Tobin Y (eds.) (2000). *Between grammar and lexicon*. Amsterdam: John Benjamins.
 Crystal D (1997). *The Cambridge encyclopedia of language* (2nd edn.). Cambridge: Cambridge University Press.
 Sapir E (1921). *Language*. New York: Harcourt, Brace & World.

H

Holism, Semantic and Epistemic

R Vallée, Université de Moncton, Shippagan, Canada

© 2006 Elsevier Ltd. All rights reserved.

According to Epistemic Holism, one does not test hypotheses – individual sentences – one at a time; rather the process of testing hypotheses confronts the whole of science with reality. According to Semantic Holism, the meaning of words and sentences in a language depends on the meaning of each and every word/sentence in that language. The former is widely accepted, whereas the latter is very controversial. Both forms of holism are tied together by the idea that the meaning of sentences is given by their truth conditions.

Epistemic Holism

Scientific theories confront reality via hypotheses, and *prima facie*, hypotheses are tested one at a time. Carl Gustav Hempel (1905–1997) and Willard van Orman Quine (1908–2000) contended that this view was an oversimplification and argued for what is now known as the Duhem/Quine thesis or Epistemic Holism (EH): “Science faces the tribunal of experience not sentence by sentence but as a corporate body: the whole of science” (Quine, 1986: 620).

Science is not compartmentalized. For example, observations made by telescopes are backed by astronomy, physics, optics, and the mathematical theories used in all these theories. The argument behind EH is simple. Assume a logical empiricist’s theory of meaning according to which the meaning of a sentence is given by its confirmation or verification conditions. Consider now any theoretical sentence. Such a sentence cannot be reduced to one single sentence concerning experience and connecting it directly to the world. So, it has no confirmation conditions as a single, independent item. It is connected to reality through a net of sentences, some of which get their truth conditions straight from facts with no regard for other sentences, the so-called observation sentences. It follows that a hypothesis is just a sensor of

the whole science, which confronts reality and is affected as a whole through the hypothesis. Hempel (1950) wrote:

“In order to understand ‘the meaning’ of a hypothesis within an empiricist language, we have to know not merely what observation sentences it entails alone or in conjunction with subsidiary hypothesis, but also what other, non observational, empirical sentences are entailed by it, and for what other hypotheses the given one would be confirmatory or disconfirmatory. In other words, the cognitive meaning of a statement in an empirical language is reflected in the totality of its logical relationships to all other statements in that language, and not to the observation sentences alone.”

In that conception, there is no clear-cut distinction between changing the theory and changing the meaning of sentences. When a hypothesis is proven false, one can alter either other parts of the theory, the so-called auxiliary hypothesis, or the meaning of the sentences of the theory. EH is sometimes offered as a reason for Semantic Holism: EH and the empiricist theory of meaning lead to SH. If the meaning of a sentence is determined by its verification conditions and if they are holistic, the meaning of a sentence is also determined holistically. EH and Semantic Holism are independent views on different topics: scientific theories and language. However, they dovetail.

Semantic Holism

Although EH is sentence oriented, Semantic Holism (SH) covers both sentential and subsentential items. There are different versions of SH, but its basic tenet is that the meaning of a sentence/word is determined by, depends on, or is constituted by, its relation to *all* the sentences/words in a language.

The Argument from Compositionality and the Context Principle

SH is usually not advocated in and of itself. EH can motivate it. It can also be introduced as a consequence of two independently motivated, nonepistemic

linguistic principles and a truth-based semantics. It is based on two principles:

1. *Principle of Compositionality*: The meaning of a sentence is a function of the meaning of its constituents and their syntactic relationship
2. *Context Principle*: “Only in the context of a sentence does a word have meaning” (Frege, 1978: x).

To give the meaning of a word, one must give the meaning of all sentences in which it occurs; that is, the truth conditions of all these sentences. Davidson (1967: 22) followed the argument to its obvious conclusion that “only in the context of a language does a sentence (and therefore a word) have meaning.” The meaning of a word or a sentence is determined by the language to which it belongs.

SH opposes ‘Semantic Atomism,’ which holds that the meaning of a sentence/word is *independent* from the meaning of *all* the other sentences/words in a language. Semantic Atomism proposes either the referent of a word or an abstract entity, like a Fregean sense, as the meaning of a word. From then on, meanings can be piled up.

SH also opposes ‘*Semantic Molecularism*,’ which holds that the meaning of a sentence/word is determined by, depends on, or is constituted by its relation to *some* of the sentences/words in a language.

Both the Principle of Compositionality (PC) and Context Principle (CP) are *a priori*, methodological principles in theoretical semantics. Semantic Holism *per se* is not methodological in nature. Following from two *a priori* principles, it is also *a priori*. It emphasizes that the words/sentences would have no meaning outside that whole and would have a different meaning in a different whole. The PC mentions syntax and applies to complex expressions. SH applies to both simple and complex, structured expressions as well. It is also very general and is not tied to a specific notion of meaning, view on human language, or program in semantics. It simply requires sentences and truth conditions, as opposed to words and referents, to be basic in semantics. SH is also a metaphysical view on language and meaning (see Fodor and Lepore, 1992). In that respect, it has no clear purchase value in theoretical or applied linguistics. On the other hand, as long as it is not incompatible with empirical research, there is no point in rejecting (or in accepting) it.

One can argue for SH in Conceptual Role Semantics (Block, 1986). That conception of semantics is grounded on the functionalist theory of the mind, which takes thought contents to be functionally defined and the meaning of sentences/words to reflect thought contents (see *Mentalese*).

According to Conceptual Role Semantics, the meaning of a token of a linguistic expression is constituted by its inferential properties.

In this view, SH does not follow from methodological principles, but is a plain assertion about meaning. It also concerns tokens, not types, and in that respect, it is an individualistic doctrine. This version of SH is arguably a metaphysical approach to meaning. It is also designed to fit a functionalist view of mental states. Block, the main proponent of Conceptual Role Semantic (CRS), was careful to mention that CRS is not a theory, but a framework for developing semantic theories.

Problems with Semantic Holism

SH raises vexing philosophical issues. Does SH reflect substantial properties of meaning? Does it have an ontological scope? If human languages are holistic, is being holistic an essential property of human languages? Or does SH just echo our way of capturing meaning in a theory, being an artifact produced by our way of theorizing/knowing the semantics of natural languages? SH has *prima facie* implausible consequences for linguistics.

Learning Problem If SH is true, then one cannot learn the meaning of a lexical item without learning a whole language. It also implies that, to understand the meaning of one single word/sentence a speaker utters, one must be able to understand all words/sentences that speaker can utter. To master the meaning of one single word, one must master a whole language. That makes learning and understanding a language an ‘all or nothing’ affair, and it implies that learning/mastering part of a language does not make sense. Friends of holism reply that “in so far as we take the ‘organic’ nature of language seriously, we cannot accurately describe the first steps toward its conquest as learning part of a language; rather, it is a matter of partly learning” (Davidson, 1965: 7). Some (Dressner, 2002) have argued that this fits what we know about the psychology of language learning. Holism apparently opposes the PC – the meaning of one sentence cannot be determined in the way described by the PC because the whole language is involved. However, it is arguable that there is no problem in that quarter. The PC presupposes that the meaning of an unstructured item is determined, and a functional relation is applied to these items. Being determined by and being a function of are two different relationships. Insofar as determination is concerned, *prima facie*, both holism and the PC can be true.

Instability Problem SH also implies that, by adding terms to the language or by altering the meaning of one single word/sentence in a language, one alters the meaning of all words/sentences in that language. As a consequence, the language of a speaker changes constantly because words/sentences are introduced and/or the meaning of lexical items is altered. In addition, two speakers do not speak the same language or use expressions having the same meaning, because the meaning of the words/sentences of a speaker differs from the meaning of the words/sentences of another speaker insofar as they do not share the same vocabulary or produce the same sentences.

Disagreement Problem If two tokens of the same type differ in meaning, then speakers cannot disagree or agree, for that matter, on a sentence. The fact that you accept and that I reject 'Snow is white' does not show that we disagree on anything; the fact that we both accept the same sentence does not show that we agree on anything. The reason is the same in both cases: The tokens have different meaning in the two idiolects, and agreement or disagreement requires sameness of meaning.

SH can be escaped if truth and sentences are not treated as basic in semantic. It can also be avoided if the PC or the CP is rejected. The PC draws wide agreement. The CP is more controversial, and Atomism rejects it. SH inherits certain indeterminacy from the premises leading to it. What do 'is determined by,' 'depends on,' 'is constituted by,' and 'meaning' mean exactly in SH? These terms are not designed to be read in a mathematical way, and they remain controversial. How can sound, *a priori* methodological principles have dramatic metaphysical consequences? Those questions do not make the problem disappear, but prove how opaque it is. The Conceptual Role Semantic version of Semantic Holism can be rejected by refusing CRS.

Semantic Holism and the Philosophy of Mind

Meaning is connected to thought content. Replace the relevant words in the above principles, and you end up with the idea that thought contents have a compositional aspect and that a single thought content is determined by all of one's thought contents. Changing one belief also implies altering the web of beliefs, and hence each belief – this is the instability problem raising its head in the philosophy of mind – and to know one single thought content of a speaker, one

must know all of his other thought contents. Fodor and Lepore (1992) drew attention to the fact that, if SH is true and if meaning is echoed in thought contents, then we cannot share a thought unless we share all of our thought contents, and trying to state psychological laws is a futile project, doomed from the start.

Conclusion

Wherever one looks, one finds only implausible consequences of SH. As a consequence, it never served as the basis for a detailed research program in semantics. However, it did bring the analytic philosopher one step closer to the Hermeneutic tradition. Hermeneutics is thoroughly holistic.

See also: Compositionality: Philosophical Aspects; Context Principle; Empiricism; Mentalese; Semantic Value; Thought and Language: Philosophical Aspects; Truth Conditional Semantics and Meaning; Verificationism.

Bibliography

- Block N (1986). 'Advertisement for a semantics for psychology.' *Midwest Studies in Philosophy*, 615–678.
- Davidson D (1965). 'Theories of meaning and learnable languages.' In Davidson D (ed.) (1980). *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
- Davidson D (1967). 'Truth and meaning.' In Davidson D (ed.) (1984). *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
- Davidson D (1984). 'Truth and meaning.' In Davidson D (ed.) *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
- Dressner E (2002). 'Holism, language acquisition, and algebraic logic.' *Linguistics and Philosophy* 25, 419–452.
- Fodor J & Lepore E (1992). *Holism. A shopper's guide*. Oxford: Basil Blackwell.
- Frege G (1978). *The foundations of arithmetic*. Oxford: Basil Blackwell.
- Hempel C G (1950). 'The empiricist criterion of meaning.' In Ayer A J (ed.) *Logical positivism*. The Free Press. 108–129.
- Okasha S (2000). 'Holism about meaning and about evidence: in defence of WV Quine.' *Erkenntnis* 52, 39–61.
- Quine W V O (1980). 'Two dogmas of empiricism.' In Quine W V O (ed.) *From a logical point of view*. Cambridge, MA: Harvard University Press.
- Quine W V O (1986). 'Reply to Vuillemin.' Hahn L E & Schilpp P A (eds.) *The philosophy of W. V. Quine*. La Salle: Open Court.

This page intentionally left blank

Ideational Theories of Meaning

E J Lowe, University of Durham, Durham, UK

© 2006 Elsevier Ltd. All rights reserved.

Ideational theories of meaning are commonly attributed to 17th- and 18th-century empiricist philosophers such as Thomas Hobbes, John Locke, George Berkeley, and David Hume, and received severe criticism from 20th-century philosophers of language, notably Gottlob Frege and Ludwig Wittgenstein. Unfortunately, the work of the earlier philosophers was seriously misunderstood by their later critics, and much of the criticism was misdirected. In fact, it is highly debatable whether the empiricist philosophers in question were offering theories of *meaning* in anything like the sense that the phrase ‘theory of meaning’ would now be understood (see Hacking, 1975, chapter 5).

Locke devoted an entire book (Book III, ‘Of Words’) of his greatest work, *An Essay Concerning Human Understanding* (Locke, 1975) to the topic of language, focusing on the communicative function of language, the advantages of language, and the abuses of language. According to Locke, the main purpose of language is to serve as a medium for the communication of thought from one thinker to another. Locke, like many of his contemporaries, favored an ideational view of thought – that is, he believed that thinking is essentially an exercise of the faculty of imagination and that thoughts consist in sequences of *ideas* in the minds of thinkers. *Idea* was a term of art in early modern philosophy, as ubiquitous then as the term *concept* is in present-day philosophical writing and playing a partly similar role, to denote a component or ingredient of the contents of thoughts. However, the empiricist philosophers used it equally to denote a component or ingredient of the contents of *sensory experiences*, reflecting the tight connection they presumed to obtain between thought and perception. Locke famously claimed that words “in their primary or immediate signification stand for nothing but the ideas in the mind of him that uses them” (Locke, 1975: 405) – one of the most misunderstood claims in the history of philosophy. Modern readers of

Locke are apt to interpret him as claiming that the *meaning* of a word is an idea in the mind of the speaker, which suggests that he adopted a thoroughly subjectivist – indeed, almost solipsistic – theory of meaning. But such an interpretation mistakenly conflates *signification*, as Locke uses this term, with *meaning* in the semantic sense. Locke is not claiming that words *refer to* or *denote* ideas in the mind of the speaker, but simply that speakers primarily use words to *express* the contents of their own thoughts, for the purpose of communicating those thoughts to others. Locke could happily concede that ‘dog’ in my mouth refers to *dogs*, not to my *idea* of a dog. His point is merely that when I assert, for example, ‘Dogs bark,’ I am expressing a thought that I have concerning dogs – a thought which is about dogs by virtue of containing as one of its ingredients my *idea* of a dog.

To clarify this matter, it is useful to distinguish between three quite different kinds of relations: *semantic* relations, *cognitive* relations, and *expressive* relations (see Lowe, 1995: 145 and Figure 1).

Semantic relations are *word-to-world* relations, such as the reference relation between the word ‘dog’ and dogs. Cognitive relations are *thought-to-world* relations, such as the intentional relation between my idea of a dog and dogs, by virtue of which the former is ‘about’ the latter, or has the latter as its ‘intentional object.’ Expressive relations are *word-to-thought* relations, such as the signification relation (in Locke’s sense) between the word ‘dog’ as used by me and my idea of a dog. In these terms, modern critics of Locke and like-minded early-modern empiricist philosophers may be accused of misconstruing their account of linguistic signification as a theory of semantic relations, when in fact it is a theory of expressive relations. Locke’s primary interest lies not in semantics but in the nature of thought and its relation to language, that is, in cognition and expression. Of course, given an account of word-to-thought (expressive) relations and an account of thought-to-world (cognitive) relations, it is possible to construct an account of word-to-world (semantic) relations, although it seems that Locke himself was not much interested in doing this in any detail. Such

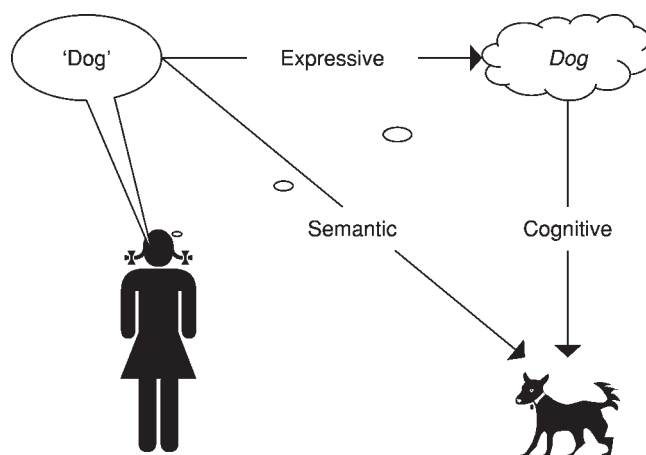


Figure 1 Locke's dog-legged semantic theory.

an account of word-to-world relations will be, in Simon Blackburn's vivid phrase, a 'dog-legged' semantic theory (Blackburn, 1984: 40), because it takes such relations to be the product of two other kinds of relations, cognitive and expressive. There may be problems with semantic theories of this type, but they will only be obscured by misunderstanding the Lockean theory of linguistic signification as *itself* being a theory of semantic, as opposed to expressive, relations. In saying that words are *signs* of ideas in the minds of speakers, Locke means that they are *indicators* of those ideas, which speakers can exploit to enable an audience to gain knowledge of what they are thinking (see Ott, 2004). Locke presumes that signification of this kind is artificial – the product of human invention – rather than natural, but that it is otherwise comparable to the indicator relation between dark clouds and impending rain, the former constituting evidence for the latter.

To evaluate the Lockean approach to language, we need to probe a little more deeply into his account of thought and ideas. Locke divides ideas into *simple* ideas of sensation and reflection, and *complex* ideas that are compounded by the mind out of those simple ideas. By 'reflection' Locke means what would now be called 'introspection.' Examples of simple ideas of sensation would be our ideas of colors, taste, and sounds, while examples of simple ideas of reflection would be our ideas of basic mental activities, such as thinking, desiring, and willing. According to Locke, many of our complex ideas are acquired by the mental process of *abstraction*, or what we might now call 'selective attention,' when we notice that certain types of simple ideas regularly accompany each other in our experience. For example, our complex idea of an *apple* will include various simple ideas of shape, size, color, and taste which we find that we regularly experience in conjunction with one another. It is a matter for dispute among Locke scholars whether or

not he conceived of sensory ideas as mental *images*, and the textual evidence is far from conclusive (see Lowe, 1995: 35–47). It is much clearer that Berkeley, writing a little later, held an imagistic view of ideas and, perhaps wrongly, construed Locke as holding one too. Berkeley famously criticized Locke's theory of abstraction as being incoherent, but the cogency of the criticism seems to depend upon the interpretation of Locke's view of ideas as being imagistic. (Berkeley urged that Lockean abstract ideas must lack determinacy of content in a way that seems problematic only on the assumption that such ideas are, or are relevantly like, images: see Lowe, 1995: 158–161.) Setting aside the controversy over whether or not Locke was an imagist, the essential features of his ideational conception of thought reduce to the following. First, Locke is clearly committed to a strong version of the doctrine that thought is independent of language. Indeed, his central aim in discussing language is to show how easily language can lead us astray if we fail to notice the differences between it and thought. Second, he clearly believes that thinking, involving as he takes it an exercise of the faculty of imagination, is intimately related to and, with respect to its content, ultimately wholly indebted to *perceptual experience*, both sensory and introspective.

Modern critics of ideationism are not only apt to misconstrue the ideational account of linguistic signification as being a semantic, as opposed to an expressive, theory, but also to oversimplify the object of their criticism. For instance, it is sometimes lampooned as maintaining that *every* word in a speaker's mouth signifies a discrete idea in the speaker's mind, including words such as 'not' and 'but,' which Locke himself classifies as 'particles' and to which he devotes a chapter in the *Essay*. It is easy enough to poke fun at the suggestion that 'not' signifies an idea of 'notness,' in the way that 'red' supposedly signifies an idea of redness. But Locke himself suggested nothing so

preposterous, contending instead that negative particles are used to convey a speaker's mental act or attitude of *denial* with respect to a certain thought-content (Locke, 1975: 471). Berkeley's version of ideationism was still more sophisticated, recognizing many uses of language other than simply to express the speaker's ideas – for example, to invoke an emotive response in an audience – and allowed, too, that often we think 'in words' rather than exclusively 'in ideas' (see Olscamp, 1970: 130–153).

Undoubtedly, the disfavor into which ideationism fell during the 20th century was largely due to the conviction that it rendered linguistic meaning excessively subjective. Frege's attack on it was integral to his more general onslaught on psychologism, which he saw as a dire threat to the objectivity of logic and mathematics. This is why he is at pains to distinguish sharply between 'ideas,' which he regards as purely subjective psychological entities, and 'senses' of expressions, which he regards as mind-independent and intersubjectively graspable abstract objects (see Frege, 1960). Wittgenstein is equally antagonistic toward ideationism, which is a prime target of his famous 'private language argument' (see Wittgenstein, 1958: 94–96). Here again the complaint is that ideas are unsuited by their irredeemably subjective and private character to be recruited for a workable account of intersubjective linguistic meaning and communication, and that ideationism unavoidably degenerates into some form of scepticism or relativism.

To the extent that criticisms focusing on the privacy of ideas construe ideationism as postulating ideas as the *meanings* of words, they are misplaced for the reasons explained above. Even so, it is fair to ask of the ideationist how words can serve to *communicate* ideas, given the privacy of the latter – a privacy that Locke himself acknowledged and emphasized. Indeed, for Locke, it is precisely *because* ideas are private – 'invisible, and hidden from others' – that language, in the form of 'external sensible signs,' is needed to 'lay them before the view of others' (Locke, 1975: 405). One might suppose it to be a fatal difficulty for ideationism that no one has direct access to the ideas of another speaker and so is never in a position to tell whether or not the idea that he or she associates with a given word resembles the idea that is associated with it in the mind of someone else. However, Locke himself was fully cognizant of this seeming difficulty and was not at all disconcerted by it. It was he, indeed, who first drew attention to the notorious puzzle now known as the 'inverted spectrum' problem – the question of how I can tell whether the way in which blue things look to me might not be how yellow things look to someone else, and vice versa

(see Locke, 1975: 389). Locke's answer is that it simply doesn't matter, for the purpose of the successful communication of thoughts between people concerning blue or yellow things. However, one might agree with Locke about this while failing to see how he was in a position to say it himself, given his commitment to an ideational theory of linguistic signification. For one might suppose that such a theory takes success in communication to consist in the *replication* in the hearer's mind of ideas which the speaker associates with the words that he or she utters. But there is no reason to tie ideationism to such a doctrine, nor any evidence that ideationists such as Locke espoused it. Ideationism is at most committed to the thesis that in successful communication of the speaker's thoughts to a hearer, ideas are evoked in the hearer's mind which *correspond* to those in the speaker's mind, in a sense of 'correspondence' which does not imply resemblance or replication. That such a correspondence obtains is subject to intersubjective confirmation without imposing upon the persons concerned the impossible burden of comparing each other's ideas, and it may be taken to be set up through the social processes of language teaching and learning (see Lowe, 1996: 172–177).

See also: Empiricism; Private Language Argument; Thought and Language: Philosophical Aspects.

Bibliography

- Ayers M R (1991). *Locke*. London/New York: Routledge.
- Berkeley G (1949). *The works of George Berkeley, Bishop of Cloyne*. Jessop T E & Luce A A (eds.). London: Nelson.
- Blackburn S (1984). *Spreading the word: groundings in the philosophy of language*. Oxford: Clarendon Press.
- Frege G (1960). 'On sense and reference.' In Geach P & Black M (eds.) *Translations from the philosophical writings of Gottlob Frege*, 2nd edn. Oxford: Blackwell.
- Hacking I (1975). *Why does language matter to philosophy?* Cambridge: Cambridge University Press.
- Locke J (1975). *An essay concerning human understanding*. Nidditch P H (ed.). Oxford: Clarendon Press.
- Lowe E J (1995). *Locke on human understanding*. London/New York: Routledge.
- Lowe E J (1996). *Subjects of experience*. Cambridge: Cambridge University Press.
- Lowe E J (2005). *Locke*. London/New York: Routledge.
- Olscamp P J (1970). *The moral philosophy of George Berkeley*. The Hague: Martinus Nijhoff.
- Ott W R (2004). *Locke's philosophy of language*. Cambridge: Cambridge University Press.
- Wittgenstein L (1958). *Philosophical investigations* (2nd edn.). Anscombe G E M (trans.). Oxford: Blackwell.

Identity and Sameness: Philosophical Aspects

H W Noonan, University of Nottingham,
Nottingham, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

'Identity' and 'sameness' mean the same; their meanings are identical. However, they have more than one meaning. A distinction is customarily drawn between **qualitative** and **numerical** identity and sameness. Things with qualitative identity share properties; so things can be more or less qualitatively identical. Poodles and Great Danes are qualitatively identical because they share the property of being a dog, but two poodles will (very likely) have greater qualitative identity. Numerical identity, however, is not a matter of more or less. Numerical identity requires absolute qualitative identity and can hold only between a thing and itself. Its name implies the controversial view that it is the only identity relation in accordance with which we can properly count (or number) things or at least that so counting has a special status: *x* and *y* are to be counted as one just in case *x* is numerically identical with *y*.

Numerical identity is our topic. It is at the center of various philosophical puzzles but seems to many to be utterly unproblematic – for what could be less problematic than the notion just characterized: that of the relation that everything has to itself and to nothing else? A possible resolution of this apparent paradox (Lewis, 1986) is that there are indeed genuine puzzles stated in the language of numerical identity, but they are not genuinely puzzles about identity, as they can be restated not using it. Philosophical debates about identity of which this is not true are just confusions.

The Logic of Identity

Numerical identity can be characterized, as just done, as the relation everything has to itself and to nothing else. But this is circular, because 'nothing else' just means 'no numerically non-identical thing.' It can be defined, equally circularly, as the smallest equivalence relation (an equivalence relation being one that is reflexive, symmetric, and transitive). Other circular definitions are available. Usually it is explained as the equivalence relation satisfying Leibniz's Law, the principle of the indiscernibility of identicals, that if *x* is identical with *y* then everything true of *x* is true of *y*. Intuitively this is right, but only satisfactory if 'what is true of *x*' is understood to include 'being identical with *x*'; otherwise, it is too weak to serve uncontroversially as a unique characterization of

identity. Circularity is thus not avoided. The converse of Leibniz's Law, the principle of the indiscernibility of identicals, that if everything true of *x* is true of *y*, *x* is identical with *y*, is correspondingly trivial if 'what is true of *x*' is so understood, but with 'what is true of *x*' restricted, e.g., to qualitative, non-relational properties, becomes philosophically controversial.

Leibniz's Law itself has been subject to controversy in the sense that the correct explanation of apparent counterexamples has been debated. It is informative to be told that Hesperus is Phosphorus, but not that Hesperus is Hesperus (Frege, 1969); Giordano was so-called because of his size, Barabelli was not (Quine, 1963). Are these not cases in which something is true of *x* (Hesperus, Giordano) but not of the identical *y*? No. The cases only illustrate the need for distinguishing Leibniz's Law from the principle of substitutivity: if '*x*' and '*y*' are co-designators they are substitutable everywhere without change of truth-value. The cases are not counterexamples to Leibniz's Law, but how best to explain such failures of substitutivity is debatable.

Relative and Absolute Identity

The view that characterizes identity as the equivalence relation that everything has to itself and nothing else and that satisfies Leibniz's Law may be called the classical view. These formal properties ensure that within any theory expressible by means of a fixed stock of one- or many-place predicates, quantifiers and truth-functional connectives any two predicates that can be regarded as expressing identity will be extensionally equivalent. They do not, however, ensure that a two-place predicate does express identity within a particular theory, for it may simply be that the descriptive resources of the theory are insufficiently rich to distinguish items between which the equivalence relation it expresses holds (Geach, 1972). For example, 'has the same income as' will satisfy these conditions in a theory whose descriptive resources do not enable one to distinguish persons with the same income (by not containing, e.g., the predicates 'is male,' 'is female').

Following Geach, call a two-place predicate with these formal properties in some theory an 'I-predicate' relative to that theory. Relative to a richer theory the same predicate, identically interpreted, may not be an I-predicate. If so it will not, and did not even in the poorer theory, express identity.

Quine has suggested that when a predicate is an I-predicate in a theory only because the language is

limited, its sentences always can be reinterpreted so that the I-predicate in the newly interpreted theory does express identity. Each sentence will retain its truth-conditions, but the references of its subsentential parts will be different. Thus, Quine suggests, in a language that cannot distinguish possessors of the same income the predicates may be reinterpreted so that the predicate that previously expressed **having the same income** expresses identity. Correspondingly, there will be a shift in the ontology of the theory (from people to income groups) (Quine, 1963).

This situation provides the basis for one philosophical controversy about identity, whether identity is relative. Geach maintains that as no criterion can be given by which a predicate expressing an I-predicate may be determined to express, not merely indiscernibility relative to the language, but absolute indiscernibility, we should jettison the classical notion of identity (Geach, 1991). Others, following Quine, regard classical identity as an always-legitimate interpretation of an I-predicate and indispensable to our understanding of quantification (Quine, 1963; Dummett, 1991; Hawthorne, 2003).

Another, related, question, is whether numerical identity deserves its name. The procedure we use in counting is as follows: assign '1' to an object *x* and to whatever bears *R* to it and to nothing else, assign '2' to an object *y* to which '1' has not been assigned, to whatever bears *R* to it and to nothing else. . . . The number arrived at will be the number of objects in the domain counting by *R*. If *R* is an equivalence relation this procedure will yield a unique result. If *R* is not numerical identity a larger count will be possible using numerical identity, but if *R* is numerical identity no larger count will be possible. Thus, things related by numerical identity have to be counted as one, whereas things related by weaker equivalence relations can be, but need not be, counted as one. So if numerical identity, classically characterized, is ever an intelligible interpretation of an I-predicate, it deserves its name.

Criteria of Identity

Leibniz's Law characterizes identity but provides no means of recognizing it – we cannot ever determine that *x* is *y* by first of all checking that everything true of *x* is true of *y*. Thus, we apply **criteria** of identity, standards by which identity is to be judged. Different kinds of thing have different criteria of identity and criteria of identity themselves have different logical forms. One distinction is between one-level and two-level criteria of identity (Williamson, 1990) for objects. The criterion of identity for numbers is two-level. The number of *F*s is the number of

*G*s if and only if a one-one correlation exists between the concept *F* and the concept *G*. Numbers are the objects for which the criterion is stated but the condition is given as a relation not on those objects but on concepts (Frege, 1950). By contrast, the criterion of identity for sets given by the Axiom of Extensionality (sets are the same if they have the same members) and Davidson's (1980) criterion of event identity (events are the same if they have the same causes and effects) are one-level, the condition is given as a relation on the very objects for which the criterion is stated. Not all criteria of identity can be two-level. But it does not follow that the only legitimate notion of a criterion of identity remaining is that of a one-level criterion. Another option (Dummett, 1981) is to deny that a criterion of identity must be regarded as a criterion of identity for a type of **object**. At a basic level, on this view, what a criterion of identity is a criterion for, is the truth of a statement in which **no** reference to objects occurs. Such a statement may be expressed using demonstratives, for instance, by saying, 'This is the same cat as that,' pointing first to a tail and then a head. In this the demonstratives need not be construed as referring to objects anymore than reference is made to objects in a feature placing sentence such as, 'It's cold here' or 'It's colder here than there.'

Identity over Time

Criteria of identity can be employed synchronically, as in the example just given, or diachronically, to determine identity over time. Identity over time is a controversial notion, however, because time involves change. Heraclitus argued that one could not bathe in the same river twice because new waters were ever flowing in. Hume argued that identity over time was a fiction we substitute for a collection of related objects. Such views can be seen as based on a misinterpretation of Leibniz's Law: if a thing changes something is true of it at the later time that is not true of it at the earlier, so it is not the same. The answer is that what is true of it at the later time is, say, 'being muddy at the later time,' which was always true of it; similarly, what is true of it at the earlier time, suitably expressed, remains true of it. But the question remains of how to characterize identity across time through change given that such identity exists. On one view, a persisting object is to be thought of as perduring, having different temporal parts at different times, on another it is to be thought as enduring, existing at different times in its entirety (Lewis, 1986). On the former view, temporary qualities like being muddy are genuinely qualities but are possessed by the temporal parts of the persisting object, not the

object itself. On the latter view, they are relations in disguise, relations that the persisting object stands in to different times.

Contingent Identity

Another controversy about identity is whether it can obtain contingently. Of course, there are contingently true identity statements – ‘Benjamin Franklin was the first Postmaster General of the United States’ – but Kripke (1980) argued that when the expressions flanking the identity sign were rigid designators designating the same object across all possible world (unlike ‘the first Postmaster General’) the identity statement had to be necessarily true, if true at all. Relatedly, he argued that identity itself was a necessary relation.

But some examples suggest that this conclusion is too sweeping – that even identity statements containing rigid designators may be, in a sense, contingently true. Consider a statue, Goliath, and the clay, Lump, from which it is composed (Gibbard, 1975). Imagine that Lump and Goliath coincide in their spatio-temporal extent. It is tempting to conclude that they are identical. But they might not have been. Goliath might have been rolled into a ball and destroyed, Lump would have continued to exist. The two would have been distinct.

The debate over contingent identity is concerned with the proper analysis of such examples.

Vague Identity

Like the impossibility of contingent identity, the impossibility of vague identity appears to be a straightforward consequence of the classical concept of identity (Evans, 1978). For if *a* is only vaguely identical with *b*, something is true of it – that it is only vaguely identical with *b* – that is not true of *b*, so by Leibniz’s Law, it is not identical with *b* at all. Of course, there are vague statements of identity – ‘Princeton is Princeton Borough’ (Lewis, 1988; ‘Princeton’ is indeterminate in denotation in standard usage between the township and the borough) – but the conclusion appears to follow that such vagueness is only possible when one or both of the terms flanking the identity sign is an imprecise designator. Relatedly, it appears to follow that identity itself must be a determinate relation.

But some examples suggest that this conclusion is too sweeping – that even identity statements

containing precise designators may be, in some sense, indeterminate. Consider Everest and some precisely defined hunk of rock, ice and snow, ‘Rock,’ of which it is indeterminate whether its boundaries coincide with those of Everest. It is tempting to think that ‘Everest’ and ‘Rock’ are both precise designators (if ‘Everest’ is not is anything?) and that ‘Everest is Rock’ is nonetheless in some sense indeterminate (Tye, 2000).

The debate over vague identity is concerned with the proper analysis of such examples.

See also: Coreference: Identity and Similarity.

Bibliography

- Davidson D (1980). *Essays on actions and events*. Oxford: Clarendon Press.
- Dummett M (1981). *The interpretation of Frege’s philosophy*. Cambridge, MA: Harvard University Press.
- Dummett M (1991). ‘Does quantification involve identity?’ In Lewis H A (ed.) *Peter Geach: philosophical encounters*. Dordrecht: Kluwer Academic Publishers.
- Evans G (1978). ‘Can there be vague objects?’ *Analysis* 38, 208.
- Frege G (1950). *The foundations of arithmetic*. Austin J L (trans.). Oxford: Basil Blackwell.
- Frege G (1969). *Translations from the philosophical writings of Gottlob Frege*. Geach P & Black M (trans.). Oxford: Blackwell.
- Geach P (1972). *Logic matters*. Oxford: Basil Blackwell.
- Geach P (1991). ‘Replies.’ In Lewis H A (ed.) *Peter Geach: philosophical encounters*. Dordrecht: Kluwer Academic Publishers.
- Gibbard A (1975). ‘Contingent identity.’ *Journal of Philosophical Logic* 4, 187–221.
- Hawthorne J (2003). ‘Identity.’ In Loux M J & Zimmerman D W (eds.) *The Oxford handbook of metaphysics*. Oxford: Oxford University Press.
- Kripke S (1980). *Naming and necessity*. Oxford: Basil Blackwell.
- Lewis D (1988). ‘Vague identity: Evans misunderstood.’ *Analysis* 48, 128–130.
- Lewis D (1986). *On the plurality of worlds*. Oxford: Basil Blackwell.
- Quine W V O (1963). *From a logical point of view*. New York: Harper and Row.
- Tye M (2000). ‘Vagueness and reality.’ *Philosophical Topics* 28, 195–209.
- Williamson T (1990). *Identity and discrimination*. Oxford: Basil Blackwell.

Immunity to Error through Misidentification

K Romdenh-Romluc, University of Nottingham, Nottingham, UK

© 2006 Elsevier Ltd. All rights reserved.

Suppose I overhear a conversation and so come to believe the fact in (1).

- (1) Ernie owns a poodle.

Two simple errors I may make are these: I may be mistaken because Ernie owns a greyhound. If so, I have predicated the wrong thing of Ernie. Or I may be mistaken because it is Bert who owns the poodle. In this case I have misidentified the poodle's owner. The latter is called an error of misidentification. To say that an utterance is immune to error through misidentification (IEM) is to say that this sort of mistake has been ruled out. If I see a dog in front of me and so form the belief in (2), I cannot be wrong about *which* dog my judgment is about (although I can make other sorts of mistakes – perhaps the dog is not a highland terrier, or perhaps I am hallucinating and there is no dog).

- (2) That dog is a highland terrier.

Whether or not an utterance about something is IEM depends on how the utterer is thinking about that thing and how he or she has come to know facts about it. The phenomenon interests philosophers because certain 'I'-utterances are IEM, and since many writers take 'I'-utterances to express first-person thoughts, it seems that this fact can shed light on first-person thinking. The next section explains how someone must think about/come to know about *x* for an utterance about *x* to be IEM. In a later section, the main positions in the debate about IEM 'I'-utterances are presented.

IEM

There are different ways for an utterance to be IEM. To come to grips with the phenomenon, we first need to understand what it is for an utterance to be *open* to this kind of error (OEM). In the first example, if the poodle belongs to Bert, I make an error of misidentification when I judge (1), because I mistake Ernie for the poodle's owner. In other words, I judge that Ernie is numerically identical with the owner of the poodle, and my judgment is false, because the person who owns the poodle is Bert. Thus we can see that an error of misidentification consists in mistakenly judging that one thing is numerically identical to another, and it is primarily judgments of numerical identity, as in (3), that are OEM.

- (3) The person who wrote *Crime and Punishment* is Dostoevsky.

Some subject-predicate judgments are also OEM. This is because they depend on judgments of numerical identity (Evans, 1982: 180). Suppose that I see a llama standing in a field and wearing a blue blanket and make the subject-predicate judgment in (4).

- (4) My sister's llama is wearing a blue blanket.

That, in turn, depends on two further judgments. One is a perceptual-demonstrative judgment about *that* llama (the one I can see), as in (5).

- (5) That llama is wearing a blue blanket.

The other is a judgment of numerical identity, as in (6).

- (6) That llama is my sister's llama.

If (6) is false, in judging (4), I will misidentify the llama that is wearing a blue blanket. Hence, (4) is OEM, because it depends on (6).

Since a judgment will be OEM if it either is or depends upon a judgment of numerical identity, it follows that a judgment will be IEM if it is not of these types. There are many different kinds of judgments that are IEM. However, discussion in the literature tends to focus on IEM subject-predicate judgments. These are perceptual-demonstrative judgments – i.e., judgments that are wholly based on direct awareness of the thing the judgment is about, as in (5).

- (5) That llama is wearing a blue blanket.

It is clear that (5) is IEM – I cannot be wrong about *which* llama is wearing a blue blanket (although my judgment can go wrong in other ways, e.g., if the blanket is not blue, there is no blanket, or there is even no llama). The example in (5) is IEM because I both identify the object of the judgment (the llama) and come to know that it is wearing a blue blanket on the basis of my seeing it in the field. As a result, (5) does not depend upon a judgment of numerical identity, and so the possibility of error through misidentification has been ruled out.

IEM 'I'-Utterances

Philosophers are primarily interested in IEM because some 'I'-utterances are IEM. It seems, e.g., impossible to utter the sentence in (7) and be mistaken about *which* person is in pain.

- (7) I am in pain.

It is widely held that 'I'-utterances express first-person thoughts. If this is so, then the fact that some

are IEM can shed light on the nature of first-person thinking, and the self. Nearly all writers accept that some 'I'-utterances are IEM (see, however, Campbell, 1999), but the source of this immunity is disputed.

What is perhaps the most influential account of IEM 'I'-utterances assimilates them to perceptual-demonstrative judgments. Those 'I'-utterances that are IEM appear to express subject-predicate judgments. When I utter (7), I seem to be saying that some object (myself) satisfies the predicate 'in pain'. We see in the first section of this article that a subject-predicate judgment will be IEM if it is wholly based on direct awareness of the thing the judgment is about. Many writers are thus led to conclude that an utterance such as (7) is IEM because it is wholly based on direct awareness of oneself – I observe myself and observe that I am in pain. These writers thereby treat 'I' as a perceptual-demonstrative comparable to terms such as 'this' and 'that' (note, however, that these terms also have uses that are not perceptual-demonstrative). But 'I' does not behave like other perceptual-demonstrative terms. I can use a term such as 'that' to refer to lots of different things; sometimes an utterance of 'that' will fail to refer to anything – e.g., if I hallucinate a llama before me, my attempts to refer to it using 'that' will fail. In contrast, it is traditionally claimed that I can only refer to myself using 'I,' and 'I' cannot fail to refer. Thus, the self-awareness that underlies 'I'-utterances that are IEM cannot be like ordinary forms of observation. First, I use a perceptual-demonstrative term such as 'that' to refer to something I currently see. 'That' can be used to refer to lots of different things, because I can see many entities. If 'I' is a perceptual-demonstrative that can be used to refer only to oneself, it follows that its use must be based on awareness one can have only of oneself. Second, perceptual-demonstratives can fail to refer because we can have illusory experiences, where we seem to perceive an object that is not in fact there. Since 'I' cannot fail to refer, the awareness that underlies its use must not allow for the possibility of complete hallucination in this way.

Various objections to the perceptual-demonstrative account of IEM 'I'-utterances have been raised. Anscombe (1981) objected that the only sort of awareness that could fit the bill would be introspection of a nonphysical self. This claim creates trouble for the view that IEM 'I'-utterances should be treated as perceptual-demonstrative judgments because theorists these days take the self to be a physical entity. A further problem was raised by Shoemaker (1968). On the perceptual-demonstrative view, I make an utterance like (9) on the basis of direct awareness

of the thing the judgment is about – myself. However, Shoemaker argued that observing a self and noticing that it is in pain only allows me to judge (8) in the same way that observing a llama and noticing that it is wearing a blue blanket allows me to judge (5).

(8) That self is in pain.

(5) That llama is wearing a blue blanket.

But (8) is clearly not equivalent to (7). First-person thoughts have a special role in bringing about action (see **Essential Indexical**). I may, e.g., see myself reflected in a shop window, notice that the person I see reflected is standing in the way of the street sweeper, and so judge (9).

(9) That person is obstructing the street sweeper.

But even if I am cooperative, I will not move out of the way unless I believe (10).

(10) I am obstructing the street sweeper.

If observing a self and noticing that it is in pain allows me only to judge (8), to judge (7), I will have to make the judgment of numerical identity (11).

(8) That self is in pain.

(7) I am in pain.

(11) I am that self.

It is unclear on what basis I could judge (11). Moreover, as we have seen, if a subject-predicate judgment depends upon a judgment of numerical identity, then it is OEM. It follows that if I judge (7) on the basis of (8) and (11), then (7) will be OEM. But (7) is IEM; thus, I cannot judge (7) on the basis of (8) and (11). Shoemaker concluded, therefore, that IEM 'I'-utterances cannot be based on self-observation.

Evans (1982) attempted to give a perceptual-demonstrative account of IEM 'I'-utterances that did not fall prey to these objections. We have seen that 'I' differs from other perceptual-demonstrative terms in that I can refer only to myself using 'I,' and 'I' cannot fail to refer. It follows that IEM 'I'-utterances must be based on a form of awareness that one can have only of oneself and that does not allow for hallucination in the same way as a sense like vision, where I can seem to see an object that is not there. Evans suggested that proprioception – the 'inner' sense of one's body, its temperature, limb position, movements, etc. – satisfies both of these conditions. Indeed, it gives rise to 'I'-utterances that are seemingly IEM. If, e.g., I feel that my legs are crossed and utter (12), I surely cannot be wrong about which person has their legs crossed.

(12) I have my legs crossed.

Evans held that all IEM 'I'-utterances are based upon proprioceptive experience of one's body.

Anscombe objected to the perceptual-demonstrative account of IEM 'I'-utterances, because it seemed that only introspection of a nonphysical self could satisfy the relevant conditions. Evans's account silenced this objection, because proprioception is awareness of one's body. It could also deal with the worry raised by Shoemaker. Shoemaker claimed that observation of a self could license only a judgment such as (8), rather than (7).

(8) That self is in pain.

(7) I am in pain.

Judgments (7) and (8) are not equivalent, because they have different implications for action. One might think that Shoemaker could make the same objection to Evans's account: proprioceptive experience allows me only to judge, e.g., (13) and not the first-person judgment (12).

(13) That body has its legs crossed.

(12) I have my legs crossed.

However, proprioceptive experience allows us to act unreflectively on the information given in that experience. Suppose that I feel a mosquito crawling on my arm; I can immediately slap my arm without thinking about doing so. In particular, I do not need to identify the arm I can feel the mosquito crawling along as *my* arm. The way in which proprioceptive awareness has an immediate bearing on my actions suggests that it is awareness that is distinctively first-personal. It follows that it can license a first-person judgment such as (12), not merely a judgment such as (13). Thus, Evans's account does not fall prey to Shoemaker's objection.

Evans's account, however, is not free from problems. It is perhaps conceivable that I could have proprioceptive experience of another's body if my brain were appropriately 'wired' to their limbs. In such a case I might judge (14).

(14) I have an itch in my left foot.

But since my proprioceptive experience comes from their body, not mine, in judging (14), I misidentify the person whose left foot is itchy. Examples such as this are famously taken to show that judgments based on proprioceptive awareness of one's body are not IEM.

Wittgenstein, who is credited with first noticing that certain 'I'-utterances are IEM, gives a radically different analysis of such utterances. To understand his account, we need to distinguish between

utterances and judgments. I have so far spoken as if they are interchangeable. In many cases this is acceptable, because utterances are often used to express judgments. Suppose I am asked a question in class and answer (15).

(15) The capital of England is London.

My utterance expresses a judgment that London is the capital of England – by uttering (15), I express my commitment to the existence of a particular country, the existence of a particular city, and the obtaining of a particular relation between them. However, not all utterances express judgments. Suppose that I utter (16) as the punch line of a joke.

(16) John's donkey is drunk.

Unlike (15), (16) does not express a judgment – I am not committed to the existence of John or his donkey. Since an error of misidentification consists in mistakenly judging that one thing is numerically identical to another, if I make an utterance that expresses no judgment, my utterance will be IEM. When I utter (16), I am not committing myself to the existence of some particular donkey, which I believe to be drunk. It thus makes no sense to suppose that I could misidentify the donkey in question.

Wittgenstein (1953) accounted for the IEM status of certain 'I'-utterances in this way. He argued that some present-tense 'I'-utterances that involve psychological predicates, such as (17), cannot be understood as expressing judgments.

(17) I feel dazed.

Instead, we should think of them as expressing one's psychological states in much the same way that groaning sometimes expresses pain. 'I'-utterances of this sort are IEM; they have this status, according to Wittgenstein, because they do not express judgments. A consequence of this view is that 'I,' as it is used in these utterances, does not refer to anything. It makes no sense to talk of a groan as having a referent. Since utterances like (17) express psychological states in the same kind of way that groaning sometimes expresses pain, 'I' as it is used in (17) does not refer to anything. Most have dismissed Wittgenstein's 'no-reference thesis,' as this view is called. It is also unclear how IEM 'I'-utterances involving *physical* predicates (assuming that there are such utterances) should be treated on the Wittgenstein picture. Suppose a particular bodily sensation prompts me to utter (12).

(12) I have my legs crossed.

It is unclear how (12) could be treated as expressing a state in the same way that groaning can be understood as expressing pain.

See also: Essential Indexical; Sense and Reference: Philosophical Aspects; Thought and Language: Philosophical Aspects.

Bibliography

- Anscombe G E M (1981). 'The first person.' In *Metaphysics and the philosophy of mind, collected papers*, vol. 2. Oxford: Blackwell. 21–36.
- Bermudez J L (1998). *The paradox of self-consciousness*. Cambridge, MA: MIT Press.
- Campbell J (1995). *Past, space and self*. Cambridge, MA: MIT Press.
- Campbell J (1999). 'Schizophrenia, the space of reasons and thinking as a motor process.' *The Monist* 82(4), 609–625.

- Cassam Q (1995). 'Introspection and bodily self-ascription.' In Bermudez J L, Marcel A & Eilan N (eds.) *The body and the self*. Cambridge, MA: MIT Press.
- Evans G (1982). *The varieties of reference*. Oxford: OUP.
- Gallagher S (2000). 'A cognitive model of immunity to error through misidentification.' In Zahavi D & Benjamin J (eds.) *Exploring the self*. Amsterdam: John Benjamins.
- O'Brien L F (1995). 'Evans on self-identification.' *Nous* 29(2), 232–247.
- Shoemaker S (1968). 'Self-reference and self awareness.' *The Journal of Philosophy* 65(19), 555–567.
- Shoemaker S (1986). 'Introspection and the self.' *Midwest Studies in Philosophy* 10, 101–120.
- Strawson P F (1959). *Individuals*. London: Methuen.
- Pryor J (1998). 'Immunity to error through misidentification.' *Philosophical Topics* 26, 271–304.
- Wittgenstein L (1953). *Philosophical investigations*. Oxford: Blackwell.
- Wittgenstein L (1958). *The blue and brown books*. Oxford: Blackwell.

Implicature

J Meibauer, Universität Mainz, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

The Basic Notions

The term '*implicature*' goes back to the philosopher Paul Grice, as laid down in his seminal article 'Logic and Conversation' (Grice, 1989), which is the published version of a part of his William James lectures held in 1967 at Harvard University. In Grice's approach, both 'what is implicated' and 'what is said' are part of speaker meaning. 'What is said' is that part of meaning that is determined by truth-conditional semantics, while 'what is implicated' is that part of meaning that cannot be captured by truth conditions and therefore belongs to pragmatics. Several types of implicature are distinguished. **Figure 1** shows the

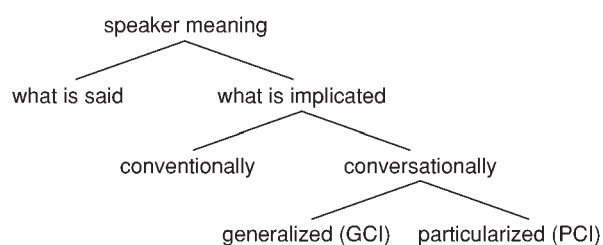


Figure 1 Gricean typology of speaker meaning.

Gricean typology of speaker meaning (cf. Levinson, 1983: 131).

The most widely accepted type of implicature is the *conversational implicature*. According to Grice, it comes in two ways, generalized conversational implicature (GCI) and particularized conversational implicature (PCI). The following example from Levinson (2000: 16–17) illustrates this distinction:

- Context, 1** Speaker A: What time is it?
 Speaker B: Some of the guests are already leaving.
 PCI: 'It must be late.'
 GCI: 'Not all of the guests are already leaving.'
- Context, 2** Speaker A: Where's John?
 Speaker B: Some of the guests are already leaving.
 PCI: 'Perhaps John has already left.'
 GCI: 'Not all of the guests are already leaving.'

Because the implicature ('... not all ...') triggered by *some* arises in both contexts, it is relatively context-independent. Relative context-independence is the most prominent property of GCIs. In addition, GCIs are normally, or even consistently, associated with certain linguistic forms. For example, if someone utters *Peter is meeting a woman this evening* it is, because of the indefinite article, standardly implicated that the woman is not his wife, close relative,

etc. (cf. Grice, 1989: 37; Hawkins, 1991). In contrast to GCIs, PCIs are highly context-dependent, and they are not consistently associated with any linguistic form.

The distinction between conversational implicatures and *conventional implicatures* draws on the observation that in coordinations like *Anna is rich but she is happy*, the truth conditions are just the truth conditions of the coordination *Anna is rich and she is happy*, with the exception of the contrastive meaning of *but*. This meaning is not truth-functional, and it is not context-dependent either; hence, there is some motivation for assuming the category of conventional implicature.

Note that there may be further types of implicature, e.g., implicatures of politeness or style that are neither conventional nor conversational (cf. Leech, 1983; Brown and Levinson, 1987).

Conversational implicatures come about by the *exploitation* (apparent flouting) or *observation* of the *cooperative principle* (CP) and a set of maxims (Grice, 1989) (see **Cooperative Principle; Maxims and Flouting**):

Cooperative Principle

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Maxim of Quantity

1. Make your contribution as informative as is required (for the current purposes of exchange).
2. Do not make your contribution more informative than is required.

Maxim of Quality

Try to make your contribution one that is true.

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Maxim of Relevance

Be relevant.

Maxim of Manner

Be perspicuous.

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly.

These maxims and submaxims are conceived as rules of rational behavior, not as ethical norms. They figure prominently in the derivation of an implicature. The basic idea of such a derivation is best illustrated with a simple dialogue. Imagine that I ask my colleague *Is Markus there?* and she answers *There is a pink Porsche behind the library building*. Understood

literally, such an answer does not make any sense. However, as I assume that my colleague is cooperative, and remembering that Markus drives a pink Porsche, I can figure out that Markus is in the library. In working out this information, I have made use of the assumption that my colleague's answer has been relevant with regard to my question. Thus, conversational implicatures display the property of *calculability*. A general scheme for the working out of a conversational implicature is given by Grice (1989: 30–31):

A man who, by (in, when) saying (or making as if to say) that *p* has implicated that *q*, may be said to have conversationally implicated that *q*, provided that (1) he is to be presumed to be observing the conversational maxims, or at least the Cooperative Principle; (2) the supposition that he is aware that, or thinks that, *q* is required in order to make his saying or making as if to say *p* (or doing so in *those* terms) consistent with this presumption; and (3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition in (2) is required.

Table 1 lists some of the most typical cases covered by the CP and the maxims. Examples for each case are given below the table. For further classical examples, see Grice (1989) and Levinson (1983). In what follows, '+>' stands for 'implicates conversationally':

- (1) War is war. +> 'There is nothing one can do about it.'
- (2) Some men were drunk. +> 'Not all of them were drunk.'
- (3a) He is a fine friend. +> 'He is not a fine friend.'
- (3b) You are the cream in my coffee. +> 'You are my best friend.'
- (4) There is life on Mars. +> 'Speaker believes that there is life on Mars.'
- (5) Speaker A: I'm out of petrol.
Speaker B: There is a garage round the corner.
+> 'The garage is open.'

Table 1 Typical cases of implicature

Maxims	Exploitation	Observation
Quantity	Tautology (1)	Scalar implicature (2)
Quality	Irony, metaphor, sarcasm (3)	Belief implicature in assertions (4)
Relevance	Implicatures due to thematic switch (5)	Bridging (6)
Manner	Implicatures due to obscurity, etc. (7)	Conjunction buttressing (8)

- (6) Speaker A: Look, that old spinster over there!
 Speaker B: Nice weather today, isn't it? +> 'No comment.'
- (7) She produced a series of noises that resembled
 "Si, mi chiamano Mimi". +> 'Her singing was
 a complete disaster.'
- (8) Anna went to the shop and bought jeans. +> 'She
 bought the jeans in the shop.'

For further illustration of the exploitation/observation dichotomy, look at (1) and (8). As to (1), tautological utterances are always true, which amounts to their being fundamentally uninformative. There is no situation where a speaker wants to tell someone that something is identical with itself. Thus, it seems that the utterer of (1) has violated the first maxim of Quality. Gricean reasoning then leads the hearer to the insight that this violation was only apparent (cf. Autenrieth, 1997). In (8), we have a simple conjunction of two sentences. If the meaning of *and* were to be the same as the meaning of the logical operator, it could not be explained that there is an additional meaning 'and then.' Grice's view is that we may identify the semantic meaning of *and* with the pure connecting operation known from logic as long as we are able to derive the additional meaning from the maxims. The observation of the fourth maxim of Manner, "Be orderly!", will do this job (cf. Posner, 1980). Both observation and exploitation are in line with the general pattern for working out an implicature.

Besides the property of calculability, conversational implicatures display the properties of variability and cancellability. *Variability* means that there are contexts where the speaker utters the same utterance, but the respective implicature does not arise. Thus, the implicature is dependent on the specific context in which it arises. (This does not exclude the notion of relative context-independency in the case of GCIs.) *Cancellability* (or defeasibility) means that it is possible to withdraw an implicature within the situation of utterance without any contradiction. For example, it is possible to utter *Some men were drunk, indeed all*. Reversely, conversational implicatures should be *reinforceable*, as Sadock (1978) proposed. Thus, it is possible to conjoin the content of an implicature with the utterance that triggers that implicature, as in *Some of the girls were reading books but not all*.

Conventional implicatures are neither calculable, nor variable, nor cancellable. However, they are said to be detachable, i.e., if the elements that trigger them are replaced, the respective implicature does not arise. By contrast, conversational implicatures are *non-detachable*, i.e., if there is an expression X' that shares meaning with expression X that triggers the implicature, the same implicature should arise. For example,

if *She is very beautiful* gives rise to an ironical implicature, then *She is a real beauty* should have the same effect (Sadock, 1978: 287). (An obvious exception to this are Manner implicatures.)

For further illustration, consider focus particles like *even*. An utterance such as *Even JOHN drives a Porsche* has the same truth conditions as the corresponding utterance without the focus particle, i.e., *John drives a Porsche*. The additional meaning of the type 'John is the least likely to drive a Porsche,' being related to a contextually given set of other individuals (e.g., Gustav, Bettina, Markus...), may be considered as a conventional implicature (cf. König, 1991), because this meaning appears to be neither truth-conditional nor context-dependent. Moreover, if *even* is replaced by another focus particle, the respective implicature is not triggered. However, if the conventional implicature is bound to the specific lexical item *even*, and for this reason is detachable, then the implicature seems to be part of the literal meaning of this lexical item. Therefore, it is difficult to distinguish between conventional implicatures on the one hand and entailments (belonging to the 'what is said') on the other hand. For this and other reasons, some researchers do not accept that there is a category of conventional implicature (cf. Bach, 1999; for a logical approach, see Potts, 2005).

Beyond Grice

The reception of the Gricean framework has been largely dominated by the wish to develop a more systematic architecture of maxims. Moreover, the Cooperative Principle has been on trial, as other aspects (e.g., logical, anthropological, cognitive, etc.) became more attractive. The prevailing tendency has been to reduce the set of maxims proposed by Grice. Three major reductive approaches have been developed: (a) the tri-heuristic approach by Levinson (2000), (b) the dual principle approach by Horn (1984), and (c) the monoprincipled approach by Sperber and Wilson (1995) and Carston (2002). These approaches are outlined in the following sections. It should be mentioned, however, that there are other important approaches that elaborate on the Gricean framework, e.g., Gazdar (1979) or Atlas (2005), as well as radical criticisms such as Davis (1998). For useful surveys, see Levinson (1983: Ch. 3) and Rolf (1994).

Presumptive Meanings: Levinson's Theory of Generalized Conversational Implicature

Levinson develops his revision of Grice's maxims from three heuristics that follow from the

anthropological need to overcome the “fundamental bottleneck in the efficiency of human communication, occasioned no doubt by absolute physiological constraints on the articulators” (Levinson, 2000: 28). Accordingly, Grice’s rationalistic CP plays no role. The heuristics are (Levinson, 2000: 31–33):

Levinson’s Heuristics

Heuristic 1: What isn’t said, isn’t.

Heuristic 2: What is simply described, is stereotypically exemplified.

Heuristic 3: What’s said in an abnormal way, isn’t normal; or Marked message indicates marked situation.

Heuristics 1 corresponds to Levinson’s Q-principle (see maxim of Quantity 1 in Grice’s framework), Heuristics 2 to Levinson’s I-principle (Grice’s maxim of Quantity 2), and Heuristics 3 to Levinson’s M-principle (Grice’s maxim of Manner 1 and 3). These three principles are said to derive GCIs. For the correspondences to Grice, and a typical example, see Table 2.

Where inconsistent implicatures arise, they are “systematically resolved by an ordered set of priorities” (Levinson, 2000: 39), among them $Q > M > I$, where ‘>’ is understood as ‘defeats inconsistency.’ Levinson (2000: 153–164) gives some examples for $Q > I$, $Q > M$, and $M > I$. An example for $Q > M$ is *It’s not unlikely that Giant Stride will win the Derby, and indeed I think it is likely*. Here, as Levinson (2000: 160) points out, the first conjunct gives rise to the M-based implicature ‘less than fully likely,’ because of the double negative *not unlikely*, while the second conjunct triggers the Q-based implicature ‘it is possible it is likely,’ because of the use of *think*, which does not entail the complement clause. In this case, the Q-implicature of the second conjunct defeats the M-implicature of the first. (However, as Traugott, 2004: 11 observes, *indeed* may serve as a M-implicature cancelling device.)

The Q-principle is defined as follows (Levinson, 2000: 76):

Q-principle

Speaker’s maxim: Do not provide a statement that is

informationally weaker than your knowledge of the world allows, unless providing an informationally stronger statement would contravene the I-principle. Specifically, select the informationally strongest paradigmatic alternate that is consistent with the facts.

Recipient’s corollary: Take it that the speaker made the strongest statement consistent with what he knows, and therefore that:

a. if the speaker asserts $A(W)$, where A is a sentence frame and W an informationally weaker expression than S , and the contrastive expressions $\langle S, W \rangle$ form a Horn scale (in the prototype case such that $A(S)$ entails $A(W)$), then one can infer that the speaker knows that the stronger statement $A(S)$ (with S substituted for W) would be false [...]

b. if the speaker asserted $A(W)$ and $A(W)$ fails to entail an embedded sentence Q , which a stronger statement $A(S)$ would entail, and $\{S, W\}$ form a contrast set, then one can infer that the speaker does not know whether Q obtains or not (i.e., $\{P(Q), P \sim (Q)\}$ read as ‘it is epistemically possible that Q and epistemically possible that not- Q ’)

The I-Principle mentioned in the Speaker’s maxim requires that a speaker should not be more informative than necessary (see below). Wherever it is possible, the speaker should build on stereotypical assumptions. In the Recipient’s corollary, two cases are distinguished, namely scalar implicature, involving Horn scales (named after Laurence Horn, see the next section) and clausal implicature, involving contrast sets.

In the case of *scalar implicatures*, we need a *Horn scale*: given a scale $\langle q, p \rangle$ with p as an informationally weak and q as an informationally strong element, the assertion of p implicates the negation of q . In such cases, the speaker is supposed to be as informative as possible, thus observing the Q-principle (or the maxim of Quantity). Therefore, the speaker could not say more than he actually did, and this means that the stronger statement does not hold. A classical example is the utterance $p = \text{Some colleagues were drunk}$ implicating $q = \text{‘Not all of them were drunk’}$. In the case of *clausal implicatures*, we need *contrast sets*. Let $\{\text{know}, \text{believe}\}$ be a contrast set. Then $p = \text{The doctor believes that the patient will not recover}$ implicates $q_1 = \text{‘The doctor may or may not know$

Table 2 Correspondences between Levinson’s Heuristics and Principles, and Grice’s Maxims

Heuristics	Principles	Grice’s Maxims	Example
Heuristic 1	Q-Principle	Quantity, 1	Q-Implicature: (a) Some colleagues were drunk. $+>$ ‘Not all of them were drunk.’ (scalar implicature) (b) The doctor believes that the patient will not recover. $+>$ ‘The doctor may or may not know that the patient will not recover.’ (clausal implicature)
Heuristic 2	I-Principle	Quantity, 2	I-Implicature: Anna turned the switch and the motor started. $+>$ ‘Anna turned the switch and then/therefore the motor started.’ (conjunction buttressing)
Heuristic 3	M-Principle	Manner, 1 and 3	M-Implicature: Bill caused the car to stop. (vs. Bill stopped the car.) $+>$ ‘He did this indirectly, not in the normal way, e.g., by use of the emergency brake.’ (periphrasis)

that the patient will not recover' (Levinson, 2000: 110). The crucial point is that clausal implicatures indicate epistemic uncertainty about the truth of the embedded sentence. Note that, because <know, believe> also form a Horn scale, there is a scalar implicature as well: in this case *p* implicates *q*₂ = 'The doctor does not know that the patient will not recover.'

Well-known Horn scales include the quantifiers <all, most, many, some>, connectives <and, or>, modals <necessarily, possibly>, <must, should, may>, adverbs <always, often, sometimes>, degree adjectives <hot, warm>, and verbs <know, believe>, <love, like>. Contrast sets include verbal doublets like {know, believe}, {realize, think}, {reveal, claim}, {predict, foresee}, and others (cf. Levinson, 2000: 111).

Now consider the I-principle (Levinson, 2000: 114–115):

I-Principle

Speaker's maxim: the maxim of Minimization. 'Say as little as necessary'; that is, produce the minimal linguistic information sufficient to achieve your communicational ends (bearing *Q* in mind).

Recipient's corollary: the Enrichment Rule. Amplify the informational content of the speaker's utterance by finding the most *specific* interpretation, up to what you judge to be the speaker's *m*-intended [= meaning-intended] point, unless the speaker has broken the maxim of Minimization by using a marked or prolix expression.

Specifically:

- a. Assume the richest temporal, causal and referential connections between described situations or events, consistent with what is taken for granted.
- b. Assume that stereotypical relations obtain between referents or events, unless this is inconsistent with (a).
- c. Avoid interpretations that multiply entities referred to (assume referential parsimony); specifically, prefer coreferential readings of reduced NPs (pronouns and zeros).
- d. Assume the existence or actuality of what a sentence is about (if that is consistent with what is taken for granted).

This principle is said to cover a whole range of implicatures: conditional perfection (9), conjunction buttressing (10), bridging (11), inference to stereotype (12), negative strengthening (13), NEG-raising (14), preferred local coreference (15), the mirror maxim (16), specialization of spatial terms (17), and possessive interpretations (18) (cf. Levinson, 2000: 117–118).

- (9) If you mow the lawn, I'll give you five dollars.
+> 'If you don't mow the lawn, I will not give you five dollars.'

- (10) Bettina wrote an encyclopedia and sold the rights to Elsevier. +> 'Bettina wrote an encyclopedia and then sold the rights to Elsevier.'
- (11) Gustav unpacked the picnic. The beer was warm. +> 'The beer was part of the picnic.'
- (12) Markus said 'Hello' to the secretary and then he smiled. +> 'Markus said "Hello" to the female secretary and then Markus smiled.'
- (13) I don't like Alice. +> 'I positively dislike Alice.'
- (14) I don't think he is reliable. +> 'I think he is not reliable.'
- (15) John came in and he sat down. +> 'John_i came in and he_i sat down.'
- (16) Harry and Sue bought a piano. +> 'They bought it together, not one each.'
- (17) The nail is in the wood. +> 'The nail is buried in the wood.'
- (18) Wendy's children +> 'those to whom she is parent'; Wendy's house +> 'the one she lives in'; 'Wendy's responsibility' +> the one falling on her; Wendy's theory +> 'the one she originated'

The M-principle is defined as follows (Levinson, 2000: 136–137):

M-principle

Speaker's maxim: Indicate an abnormal, nonstereotypical situation by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situation.

Recipient's corollary: What is said in an abnormal way indicates an abnormal situation, or marked messages indicate marked situations, specifically:

Where *S* has said *p*, containing a marked expression *M*, and there is an unmarked alternate expression *U*, with the same denotation *D*, which the speaker might have employed in the same sentence-frame instead, then where *U* would have I-implicated the stereotypical or more specific subset *d* of *D*, the marked expression *M* will implicate the complement of the denotation *d*, namely \bar{d} of *D*.

The M-principle is supposed to cover a range of cases, among them lexical doublets (19) and rival word formations (20), nominal compounds (21), litotes (22), certain genitive (23) and zero morph constructions (24), periphrasis (25), and repetition (26) (cf. Levinson, 2000: 138–153).

- (19) She was reading a tome [vs. book]. +> 'She was reading some massive, weighty volume.'
- (20) Ich nehme den Flieger [vs. das Flugzeug].
(= I take the plane [vs. the airplane]) +>
'Fliegen ist nichts Besonderes für mich.'
(= 'Flying is quite normal for me.')

- (21) This is a box for matches (vs. matchbox). +> 'This is a (nonprototypical) box specially made for containing matches.'
- (22) It took a not inconsiderable effort. +> 'It took a close-to-considerable effort.'
- (23) the picture of the child (vs. the child's picture) +> 'the picture depicts the child'
- (24) She went to the school/the church/the university (vs. to school, to church, to university, etc.) +> 'She went to the place but not necessarily to do the associated stereotypical activity.'
- (25) Bill caused the car to stop. (vs. Bill stopped the car.) +> 'He did this indirectly, not in the normal way (e.g., by using the emergency brake).'
- (26) He went to bed and slept and slept. +> 'He slept longer than usual.'

Note that only the first ('Avoid obscurity of expression') and the third ('Be brief (avoid unnecessary prolixity)') submaxims of the Gricean maxims of Manner survive in Levinson's M-principle. Levinson views the second submaxim ('Avoid ambiguity') in connection with 'generality narrowing', which is subsumed under the Q-principle (Levinson, 2000: 135). The fourth submaxim ('Be orderly') is not needed any more, because the notorious cases of 'conjunction buttressing' fall under the I-principle in Levinson's framework. Moreover, Levinson (2000: 135) notes the general cognitive status of this general semiotic principle of linearization, and he questions its status as a maxim.

It seems that many of the cases in (19)–(26) may be explained in terms of the Q- or I-principle; in other cases, it is not at all clear that we have the same denotation, as required in the Recipient's corollary of the M-principle, thus throwing into doubt whether a separate M-principle is really needed. By comparison, Horn's (1984) approach (sketched in the next section) has no separate maxim/principle of Manner. For further discussion, see Meibauer (1997) and Traugott (2004).

Obviously, the maxim of Quality and the maxim of Relevance are not maxims that figure in the derivation of GCIs. The only comment on the maxim of Quality Levinson gives is that this maxim "plays only a background role" in the derivation of GCIs; maybe he has the sincerity conditions for assertive acts in mind (Levinson, 2000: 74). Note that Grice (1989: 34) needed the maxim of Quality to derive the implicatures in the cases of irony, metaphor, and sarcasm (see *Irony*). In contrast, Levinson argues that irony and sarcasm are cases of PCIs (Levinson, 2000: 386, Fn. 2), a claim that seems somewhat premature at least when considering cases

of conventional irony and sarcasm. The maxim of Relevance is a maxim that, according to Levinson (2000: 74), derives only PCIs. However, this maxim seems to play a role when it comes to disambiguation and 'ellipsis unpacking' (Levinson, 2000: 174, 183) (see *Relevance Theory*).

In addition to the revision of the Gricean maxims just outlined, Levinson sketches a radical revision of the widely accepted Gricean view of the interaction of grammar and pragmatics according to which in language production, conversational implicatures are supposed to operate on, and follow the semantic representation of, the said (Levinson, 2000: 173). Levinson finds this view basically wrong:

Grice's account makes implicature dependent on a prior determination of 'the said.' The said in turn depends on disambiguation, indexical resolution, reference fixing, not to mention ellipsis unpacking and generality narrowing. But each of these processes, which are prerequisites to determining the proposition expressed, may themselves depend crucially on processes that look undistinguishable from implicatures. Thus, what is said seems both to determine and to be determined by implicature. Let us call this *Grice's circle*. (Levinson, 2000: 186)

According to Levinson, there are at least five phenomena that show the influence of GCIs on sentence meaning (Levinson, 2000: 172–187). First, GCIs (of the scalar type) are involved in the disambiguation of ambiguous constructions like *some cats and dogs*, for only the bracketing *[[some cats] and dogs]*, with the appropriate implicature 'some but not all cats, and dogs in general,' is appropriate in the sentence *He's an indiscriminate dog-lover; he likes some cats and dogs*. Second, the resolution of indexicals is dependent on the calculation of GCIs, e.g., *The meeting is on Thursday*. +> 'not tomorrow' (when tomorrow is Thursday). Third, reference identification often requires GCIs, e.g., *John came in and the man sat down*. +> 'The man was not identical to John.' Fourth, in ellipsis unpacking, as in simple dialogues like *Who came? – John <came>*, the missing information is constructed on the basis of Relevance and I-Implicature. Finally, there is the case of generality narrowing, e.g., if someone utters *I've eaten breakfast* +> 'I've eaten breakfast [this morning]' where the Q-principle is activated.

In order to resolve the dilemma of Grice's circle, i.e., to account for 'pragmatic intrusion,' Levinson proposes an alternative model (Levinson, 2000: 188). This model contains three pragmatic components, namely Indexical Pragmatics, Gricean Pragmatics 1, and Gricean Pragmatics 2, and two semantic components, namely Compositional Semantics and Semantic Interpretation (model-theoretic

interpretation). The output of Compositional Semantics and Indexical Pragmatics is input for Gricean Pragmatics 1. The output of Gricean Pragmatics 1 is input for Semantic Interpretation, and its output ('sentence meaning, proposition expressed') is input for Gricean Pragmatics 2, whose output is 'speaker meaning, proposition meant by the speaker.'

Whereas Indexical Pragmatics and Gricean Pragmatics 1 are presemantic pragmatic components, Gricean Pragmatics 2 is a postsemantic pragmatic component. It seems that Gricean Pragmatics 2 deals with PCIs ('indirection, irony and tropes, etc.,') whereas Gricean Pragmatics 1 deals with GCIs ('disambiguation, fixing reference, generality-narrowing, etc.,'). At the heart of Levinson's approach is his analysis of GCIs, precisely because it is here that arguments for this new model of the semantics-pragmatics interaction may be found.

Division of Pragmatic Labor: Horn's Q- and R-Principles

Central to Horn's approach to implicature is the insight that implicatures have to do with "regulating the economy of linguistic information" (Horn, 2004: 13). In contrast to Levinson, Horn (1984) assumes only two principles, the Q-principle and the R-principle:

Q-principle

Make your contribution sufficient: Say as much as you can (given R).

(Lower-bounding principle, inducing upper-bounding implicatures)

R-principle

Make your contribution necessary: Say no more than you must (given Q).

(Upper-bounding principle, inducing lower-bounding implicatures)

The Q-principle collects the Gricean maxims of Quantity 1 as well as Manner 1 and 2, while the R-Principle collects Quantity 2, Relation, and Manner 3 and 4. The maxim of Quality is considered as unreducible, as truthfulness is a precondition for satisfying the other maxims (Horn, 2004: 7).

The Q-principle aims at the maximization of content. It is a guarantee for the hearer that the content is sufficient. The hearer infers from the speaker's failure to use a more informative or briefer form that the

speaker was not in a position to do so. Scalar implicatures are a case in point. The R-principle aims at the minimization of expression, and consequently, the minimization of the speaker's effort. According to Horn, this principle holds for all indirect speech acts.

The following table, which is adapted from Horn (2004: 10), shows how the Q-principle works in the case of scalar implicatures (Table 3). The two-sided reading is the default case.

According to Horn, the conflict between the Q-principle and the R-principle may be resolved, as expressed by the following principle (Horn, 1984: 22):

The Division of Pragmatic Labor

The use of a marked (relatively complex and/or prolix) expression when a corresponding unmarked (simpler, less 'effortful') alternative expression is available tends to be interpreted as conveying a marked message (one which the unmarked alternative would not or could not have conveyed).

Levinson (1987: 73) argues that Horn mixes up two things here that properly should be distinguished, namely minimization of content on the one hand, and minimization of expression on the other. According to Levinson, splitting up the maxims of Manner in the way Horn does is mistaken, because the Manner maxims are fundamentally dependent on form, and thus related to minimization of expression.

Following Horn's original work, much research has been done on Horn scales, e.g., by Hirschberg (1991), Fretheim (1992), Matsumoto (1995), Sauerland (2004), van Rooy (2004). In this connection, three further areas of research deserve to be singled out.

First, as shown in Horn (1989: Ch. 4), there is the phenomenon of *metalinguistic negation*. For example, when uttering *It's not warm, it's hot!* the first part of the utterance gives rise to the scalar implicature 'It is not hot,' but this implicature is obviously denied in the second part of the utterance. Typically, utterances of this type have a humorous, ironical, or sarcastic flair (cf. Chapman, 1996 for an overview and Carston, 1996 and Iwata, 1998 for an echo-theoretic interpretation).

Second, there is some discussion about the exact status of the Horn scales in the lexicon, e.g., how are

Table 3 Application of the Q-Principle to scalar implicatures

<i>Statements</i>	<i>Lower bound, one-sided (what is said)</i>	<i>Upper bound, two-sided (what is implicated qua Q)</i>
a. Pat has three children	'... at least three ...'	'... exactly three ...'
b. You ate some of the cake	'... some if not all ...'	'... some but not all ...'
c. It's possible she'll win	'... at least possible ...'	'... possible but not certain ...'
d. He's a knave or a fool	'... and perhaps both ...'	'... but not both'
e. It's warm	'... at least warm ...'	'... but not hot'

elements selected for scales, how is the ordering of the elements achieved, etc. An influential approach is the one by Hirschberg (1991), who argues that there exist, in addition to lexical scales, scales that are induced pragmatically or on the basis of real-world knowledge. For example, when speaker A asks *Did you get Paul Newman's autograph?* and speaker B answers *I got Joanne Woodward's*, implicating 'not Paul Newman's,' we are dealing with a salient scale of autograph prestige <Newman, Woodward>. Consequently, Hirschberg (1991: 42) denies that there is any principled distinction between GCIs and PCIs.

Third, the economical aspect of Horn's reduction of the Gricean apparatus has recently become very attractive within *Bidirectional Optimality Theory* (cf. Blutner, 2004). This theory assumes that sentences are semantically underspecified, and therefore are in need of enrichment. A function *Gen* is assumed that determines for each common ground the set of possible enrichments. Bidirectional (i.e., taking the perspective of speaker *and* hearer) Optimality Theory then stipulates that a form-meaning pair is optimal if and only if it is taken from the set defined by *Gen*, and that there is no other pair that better fulfills the requirements of the Q- and I-principle. For an application and further discussion, see Krifka (2002).

Relevance Theory: Carston's Underdeterminacy Thesis

Relevance theory is a cognitive theory of meaning whose major claims are that semantic meaning is the result of linguistic decoding processes, whereas pragmatic meaning is the result of inferential processes constrained by one single principle, the *Principle of Relevance*, originally proposed in Sperber and Wilson (1995) (see **Relevance Theory**). However, the connection to the Gricean maxim of Relevance is rather weak, as can be seen from the following definitions (Carston, 2002; for other versions, see Wilson and Sperber, 2004):

First (Cognitive) Principle of Relevance

Human cognition is geared towards the maximization of relevance (that is, to the achievement of as many contextual (cognitive) effects as possible for as little processing effort as possible).

Second (Communicative) Principle of Relevance

Every act of ostensive communication (e.g., an utterance) communicates a presumption of its own optimal relevance.

Carston (2002) questions the standard division of labor between semantics and pragmatics and argues that pragmatics contributes much more to the

construction of explicit meaning ('what is said') than generally assumed. Her overall aim is to establish relevance theory as a theory of cognitive pragmatics. The relevance theoretic approach is, according to Carston, "to be characterized as a sub-personal-level explanatory account of a specific performance mechanism conducted at the level of representations-and-procedures" (Carston, 2002: 11).

Carston's *underdeterminacy thesis* says that linguistic meaning generally underdetermines what is said. Pragmatic inferences are not only necessary to determine implicatures, but also to fix the proposition directly expressed by an utterance. This discrepancy between the meaning encoded in linguistic expressions and the proposition expressed by the utterance of these expressions ('what is said') is illustrated by various cases (over and above the well-known cases of ambiguities and indexical resolution): missing constituents (27), unspecified scope of elements (28), underspecification or weakness of encoded conceptual content (29), overspecification or narrowness of encoded conceptual content (30):

- (27a) [Where is the book?] On the top shelf.
(= 'The book is on the top shelf.')
- (27b) Paracetamol is better. [than what?]
- (27c) This fruit is green. [which part of the fruit?]
- (28a) She didn't butter the toast in the bathroom
with a knife. [different stress changes the
information structure]
- (28b) There's nothing on TV tonight. [nothing that is
interesting for you]
- (29) I'm tired. [predicate is too weak]
- (30) Her face is oblong. [predicate is too narrow]

In all these cases, additional inferential steps are necessary to understand what the speaker intends to say.

Since linguistically encoded meanings are necessarily incomplete, pragmatics makes an essential contribution not only to the construction of *implicit meaning* but also to the construction of *explicit meaning*. In the spirit of Relevance Theory, Carston proposes a three-level model of semantic and pragmatic interpretation of linguistic expressions.

The first step involves semantic decoding of linguistic expressions. The output of the semantic decoding is an impoverished, nonpropositional semantic representation, which Carston calls logical form. It can be described as a "structured string of concepts with certain logical and causal properties" (Carston, 2002: 57) containing slots indicating where certain contextual values must be supplied. Hence, the output of the semantic decoding device is an incomplete template or scheme, open to a range of compatible propositions.

In the second step of interpretation, the hearer reconstructs the proposition intended by the speaker through pragmatic inference. Thus, pragmatic inference bridges the gap between what is linguistically expressed (incomplete conceptual schemata/logical form) and what is said (full propositional representations). For example, when a speaker utters the subsentential expression *on the top shelf* in a given context of utterance, the hearer is supposed to reconstruct the missing constituents to yield the intended proposition 'The marmalade is on the top shelf'. The pragmatic interpretation device is constrained by the First (Cognitive) Principle of Relevance, as proposed by Sperber and Wilson (1995).

Finally, there has to be a third step of interpretation, in which the hearer determines implicatures, i.e., 'what is meant.' Thus, Carston assumes that pragmatic inference is necessary for the second and third step of interpretation. In this cognitive approach, the bulk of utterance interpretation has to be done by pragmatic inference.

The pragmatic device of interpretation relies not only on linguistic information but also on additional information gained from context, perception, and world knowledge. Here, Carston essentially refers to Searle's theory of mind, especially his notion of *Background* (cf. Searle, 1980). Utterances are interpreted against a set of more or less manifest background assumptions and practices. Consider, for instance, the following five sentences: (a) *Jane opened the window*, (b) *Jane opened her book on page 56*, (c) *Jane opened the wall*, (d) *Jane opened her mouth*, (e) *The doctor opened her mouth*. Carston assumes that the encoded meaning of the English verb *open* does not vary in all five examples, although *open* receives quite different interpretations, depending on a set of background assumptions about different practices of opening. The Background is construed as a set of weakly manifest assumptions and practices in an individual's cognitive environment. Since the Background always supplies additional meaning to the interpretation of an utterance, the proposition expressed by an utterance cannot be fully determined by the meaning of its parts and the mode of their combination. Consequently, the principle of semantic compositionality does not hold for the proposition expressed, but only for the underdetermined logical form (i.e., the first step of interpretation).

As does Levinson (2000), Carston, too, argues that Grice does not account for the fact that 'what is said' is not independent from pragmatic input. However, Carston and Levinson differ in their approaches to the question of how the pragmatic intrusion problem needs to be dealt with. As shown above, Levinson develops a pragmatic subtheory of GCIs, dealing

only with the pragmatic processes involved in the elaboration of 'what is said'. By contrast, Carston favors a unitary account of all pragmatic processes, irrespective of whether they contribute to the 'what is said' or to different implicated assumptions (corresponding to Levinson's PCIs).

Carston's (2002: 377) use of the terms *explicature* and *implicature*, essentially based on Sperber and Wilson's (1995: 182) distinction between explicit and implicit assumptions/propositions, is spelled out in the following way (cf. Carston, 1988):

Explicature

An ostensibly communicated assumption that is inferentially developed from one of the incomplete conceptual representations (logical forms) encoded by the utterance.

Implicature

An ostensibly communicated assumption that is not an explicature; that is, a communicated assumption which is derived solely via processes of pragmatic inference.

The difference between explicatures and implicatures lies essentially in the way they are supplied: explicatures are developments of the logical form that they contain as a proper subpart, whereas implicatures are derived purely inferentially. In regard to these two kinds of pragmatic enrichment, the cognitive approach Carston promotes motivates the distinction between 'communicated assumptions' and the 'inferential steps' leading to them. Carston argues that explicatures are construed by means of interpretative hypotheses rather than by (generalized) implicatures.

Consider the example: *John came in and he sat down*. The preferred interpretation for the personal pronoun *he* in the second sentence is the coreferential one. Following Levinson, this interpretation results from an I-implicature. Carston argues that this implicature must be a proposition like '*He* refers to whomever *John* refers to', "a propositional form representing a hypothesis about reference assignment" (Carston, 2002: 151). She rejects the idea of reference assignment being an implicature and rather identifies it as an interpretative hypothesis like '*John* came in and he, *John*, sat down,' which is derived online and only confirmed if it meets the expectation of relevance. Carston claims that this strategy is able to resolve the dilemma of Grice's circle, for the simple reason that interpretation processes can be effected simultaneously.

Finally, the cognitive approach leads Carston to reject *conventional implicatures*; these are subsumed under the procedural elements. Relevance Theory distinguishes between concepts as constituents of mental representations, and procedures that constrain pragmatic inferences. Conventional implicatures conveyed by expressions such as *moreover* and

therefore do not contribute to the conceptual part of the utterance, but point the hearer to the kind of pragmatic processes she is supposed to perform (cf. Blakemore, 2002) (see **Meaning: Procedural and Conceptual**).

Bach (1994), who tries to defend the Gricean notion of 'what is said,' criticizes the notion of explicature and proposes instead the term *implicature* (cf. also Bach, 2001). Implicatures are either *expansions* of 'what is said,' as in *You are not going to die [from this little wound]* or *completions*, as in *Steel isn't strong enough [for what?]*. In these cases, "the resulting proposition is not identical to the proposition expressed explicitly, since part of it does not correspond to any elements of the uttered sentence"; hence Bach considers it "inaccurate to call the resulting proposition the explicit content of an utterance or an explicature" (Bach, 1994: 273).

Carston views Relevance Theory as a cognitive theory of utterance understanding that aims at the subpersonal level, where processes are fast and automatic. Thus, it should be clear that this theoretical goal differs from that pursued by Grice (cf. Saul, 2002). It must be noted, however, that arguments from psycholinguistic research are called for in order to constrain the theory.

First, it may be asked how children acquire implicatures and what roles maxims, principles, and the like play in this process. There are studies on the acquisition of irony and metaphor by Winner (1988) as well as studies on the role of Gricean principles in lexical acquisition (cf. Clark & E V, 1993, 2004). More recently, studies have been done on the acquisition of scalar implicatures, in particular dealing with the hypothesis that small children are "more logical" than older children and adults, in that they more readily accept the "some, perhaps all" – reading of the quantifier *some* (cf. Noveck, 2001; Papafragou and Musolino, 2003).

Second, there is some evidence that hearers do not first compute the literal meaning, then the nonliteral or indirect meaning, but that they arrive at the nonliteral/indirect meaning earlier or in a parallel fashion (cf. Shapiro and Murphy, 1993; Récanati, 1995; Gibbs, 2002; Giora, 2003). It is obvious that experimental research is very important for implicature and explicature theory (cf. Wilson and Sperber, 2004: 623–628).

Quality Reconsidered

In the development of neo-Gricean approaches to implicature such as Horn's and Levinson's, the Gricean maxim of Quality has been neglected. Thus, genuine pragmatic matters such as metaphor, irony,

sarcasm, lying, etc. have become largely unattractive for some implicature theorists, although metaphor had been featured as a cardinal case of maxim exploitation already early on (cf. Levinson, 1983: 147–162). Relevance Theory, on the other hand, which takes a stand on Grice as well as on neo-Gricean approaches, has developed an independent theory of irony; moreover, Carston (2002: Ch. 5), analyzes metaphors as instances of *ad hoc*-concept construction. In neither of these approaches, however, does the maxim of Quality play any role (see **Metaphor: Psychological Aspects**).

First, consider *irony*. If a speaker A utters *X is a fine friend*, referring to a person who has betrayed a secret of A's to a business rival, then the first maxim of Quality is flouted (Grice, 1989: 34). Because it is obvious that A does not believe what he says, the hearer reconstructs a related proposition, i.e., the opposite of p. The ironical implicature qualifies for the status of an implicature, because it is calculable, context-dependent, and cancellable. Note that this substitutional analysis is in contrast to the additive nature of other types of implicature. However, this approach has been criticized for several reasons: (i) The analysis cannot account for ironical questions, requests and understatements, (ii) it cannot explain the distinction between irony and metaphor, because the latter is also explained with regard to the first maxim of Quality, and (iii), it is not fine-grained enough, because it does not follow from 'He is not a fine friend' that he is not a friend at all.

The Gricean approach to irony has been most prominently attacked by relevance theorists (Sperber and Wilson, 1981; Wilson and Sperber, 1992; Sperber and Wilson, 1998). Following Sperber and Wilson, ironical utterances have four main properties: (i) They are mentioned, not used, (ii) they are **echoic** in nature, (iii) the ironical interpretation is an implicature that is derived through recognition of the echoic character of the utterance (Sperber and Wilson, 1981: 309), (iv) the ironical speaker displays a dissociative attitude towards the proposition uttered. Take the utterance *What lovely weather!* as an example. When uttered during a downpour, the speaker cannot mean the opposite, because this would be uninformative. Instead, he wants to convey that it was absurd to assume that the weather would be nice. Thus, the ironical utterance is a case of echoic mention of a previously entertained proposition. Types of echo include sarcastic repetition (31), attributed thoughts (32), norms (33) and standard expectations (34) (cf. Sperber and Wilson, 1998):

- (31) A: I'll be ready at five at the latest.
B: Sure, you'll be ready at five.

- (32) A: I'll be ready at five at the latest.
B: You mean at five tomorrow?
- (33) A: I'll be ready at five at the latest.
B: You are so punctual.
- (34) A: I'll be ready at five at the latest.
B: It's a great virtue to be on time!

Thus, the echo theory of irony does not imply that there is always an original utterance that is exactly reproduced. The echo theory is constrained in that most utterances cannot be interpreted as echoes, and echoic interpretations must contribute to the relevance of an utterance.

Several objections to this theory may be made (cf. Sperber and Wilson, 1998): (i) The notion of an echo is far too vague; it does not make sense to look for an echo in cases of conventional irony, e.g., when somebody utters *Boy, is it hot!* when it is icy cold. (ii) Because not every echoic mention is ironical, echoic mention is not sufficient to explain ironical interpretation. (iii) It is not clear why the substitution of the opposite should not be a starting point in the search for the dissociative attitude of the speaker towards the proposition. (iv) Relevance Theory cannot explain why hearers often fail to grasp the relevance of an ironical utterance.

Second, consider *metaphor*. For Carston (2002), metaphors are cases of *ad hoc* concept construction. *Ad hoc* concepts are those concepts “that are constructed pragmatically by a hearer in the process of utterance comprehension” (Carston, 2002: 322). Typical instances of *ad hoc* concepts come about via narrowing or broadening. *Narrowing* may be illustrated by utterances like *Ann is happy*, where the concept associated with *happy* in a particular context is much narrower than the encoded concept. The case of *broadening* is exemplified by utterances like *There is a rectangle of lawn at the back*, where it is very unlikely that the encoded concept of *rectangle* is communicated. Both processes are cases of constructing an *ad hoc* concept that contributes to the explicature.

If metaphors are *ad hoc* concepts, then they are part of the explicature as well. Thus, in *Mary is a bulldozer*, the logical form of *bulldozer* is associated with an *ad hoc* concept **BULLDOZER*** differing from the concept **BULLDOZER** usually encoded by this word. In this approach, metaphor isn't an implicature any more, as Grice (1989) and Levinson (1983) would have it, but an explicature.

Recall that for Horn (1984), the maxim of Quality was unreducible. Since then, its domain of application has considerably shrunk. However, it still seems to play a role when it comes to the analysis of lying, deception, insincerity, and – maybe – irony (cf. Wilson and Sperber, 2002; Meibauer, 2005). In Levinson's

(2000) approach, matters of irony, etc., are dealt with in the component called Gricean Pragmatics 2. Maybe it is there that the maxim of Quality will have a comeback. It is clear that some version of the maxim plays also a role in the definition of success conditions for assertive illocutions (see **Irony**).

Implicature and the Grammar/Pragmatics Interface

As has become clear from the sketch presented here of Levinson's and Carston's frameworks, pragmatic inferencing is powerful enough to influence semantic representations (see **Semantics–Pragmatics Boundary**). However, when it comes to pinpoint the exact relations of implicatures to illocutions on the one hand, and sentence types on the other, there still are many open questions.

First, consider implicatures *vis-à-vis* illocutions. Even if both are associated with an individual speech act, these notions refer to different entities: an additional proposition, in the case of implicature, vs. a type of act such as a promise, assertion, request etc., in the case of illocution.

An important connection between illocutions and implicatures is usually seen as obtaining in the case of *indirect speech acts* (see **Speech Acts**). According to Searle (1975), a reconstructive process that leads the hearer from the secondary illocutionary point (the 'literal' illocution) to the primary illocutionary point (the intended illocution) is similar to the scheme of reasoning that Grice proposed for conversational implicatures; step 2 of his sample derivation even includes principles of conversational cooperation (compare also the speech act schema proposed by Bach and Harnish, 1979). Accordingly, indirect speech acts have sometimes been analyzed as implicatures, for example the question *Can you close the window?*, meant as a request to close the window, a case that is related to the R-Principle as proposed by Horn (1989, 2004).

A case in point is the *rhetorical question*. Whereas Meibauer (1986) analyzes them as indirect speech acts, i.e., interrogative sentences types associated with assertive force and polar propositional content, Romero and Han (2004) analyze negative *yes/no* questions like *Doesn't John drink?* as connected with a positive epistemic implicature such as 'The speaker believes or at least expects that John drinks.' It is not clear at first sight whether such analyses are compatible; in any case, as Dascal (1994) has shown, the notions of implicature and speech act are independently motivated, and should not be confused. Thus, the question of their interrelation requires further research.

Second, consider implicatures vis-à-vis sentence types. It is widely accepted that there is a systematic connection between sentence types such as declarative, interrogative, and imperative, etc., and illocutions such as assertion, question, and request, etc.; moreover, in some approaches the existence of an intermediate category ‘sentence mood’ is assumed (cf. Sadock and Zwicky, 1985; Harnish, 1994; Reis, 1999; Sadock, 2004; Zanuttini and Portner, 2003). However, while it is conceivable that sentence types determine a certain illocutionary potential, the analogical notion of an ‘implicature potential’ has never been proposed, probably because of the authors’ concentration on lexical elements that give rise to GCIs.

However, there are several observations showing that such a concept is not totally mistaken. Consider the following examples:

- (35) Who is the professor of linguistics at
Tübingen? +> Someone is the professor of
linguistics at Tübingen.
- (36) [I gave the encyclopedia to Bettina.] You gave
the encyclopedia to WHOM?
- (37) Visit Markus and you’ll get new ideas! +> If
you visit Markus then you’ll get new ideas.
- (38a) This is good. +> This is not excellent.
(38b) Is this good? *+> Is this not excellent?

In (35), we have the case of an existential implicature that is typically bound to *wh*-interrogatives, but shows the properties of variability and cancellability. (Its classification as an existential presupposition, cf. Levinson, 1983: 184, has been abandoned, because it does not survive the negation test.) Example (36) illustrates the echo-*wh*-question. As Reis (1991) has persuasively argued on the basis of German data, these sentence types are neither ‘echo-*wh*-interrogatives’ nor *wh*-interrogatives. Instead, these utterances are regular instances of any sentence type, and their interrogative force is explained as a conversational implicature triggered by the *wh*-element (see also Reis, 1999). Another example showing that implicatures are sensitive to sentence types is the conditional imperative in (37) (cf. Davies, 1986; Clark, 1993). Finally, if elements that trigger scalar implicatures are in the scope of a question operator, the respective implicature may be blocked, as shown in (38) (the asterisk * denotes a blocked or unallowed implicature). In summary, then, there is evidence of a systematic interaction between implicatures and sentence types. The question is, then, how and where to account for this interaction.

A detailed analysis of the sentence type-implicature relation is developed in Portner and Zanuttini (2000).

They concentrate on negated *wh*-interrogatives and exclamatives in Paduan, a northern Italian dialect spoken in the city of Padua:

- (39a) Parcossa *no* ve-to anca ti!? (*wh*-interrogative)
Why NEG go-s.cl also you
‘Why aren’t you going as well!?’
- (39b) Cossa *no* ghe dise-lo! (*wh*-exclamative)
what NEG him say-s.cl
‘What things he’s telling him!’

The point is that the NEG-element has no negative force. In principle, there are two strategies for analyzing examples like (39): First, as a special type of negation, nonpropositional, expletive, or modal in character. The second strategy, as proposed in Meibauer (1990) on the basis of German data, is to assume regular negation, and to derive the modal effect from pragmatic principles. Portner and Zanuttini (2000), drawing on the latter approach, assume that exclamatives are factive. The negation particle *no* triggers a conventional implicature, which says that the lowest element from a set of alternative elements (that are possible in a contextually given scale) is true. In cases like (39a), there is an expectedness scale {less expected < more expected}, in cases like (39b), there is an unexpectedness scale {more expected < less expected}. The scales are dependent on the respective sentence type. While it is not clear (i) whether exclamatives constitute a separate sentence type at all (cf. d’Avis, 2001), (ii) why the implicatures are of the conventional type, and (iii) how the relevant scales are obtained from the context, it should be clear that such an approach paves the way for a more empirical research on the interplay of sentence types and implicatures.

Conclusions

On the basis of the foregoing sketch of three major approaches to implicature theory, we may state some of the prevailing tendencies. To begin with, there is a striving to understand implicatures in terms of economy. This is true for Levinson’s insight that implicatures help to overcome “the slowness of articulation,” as becomes clear from his slogan “inference is cheap, articulation expensive” (Levinson, 2000: 29), as well as for Horn’s appeal to the principle of least effort and Sperber and Wilson’s view on optimal relevance. Lately, recent developments in Optimality Theory have shown attempts to integrate the interplay of maxims into their frameworks.

Second, there is a tendency to reject the classic dual distinction between ‘what is said’ on the one hand, and ‘what is implicated’ on the other. Instead, a three-level approach to meaning is favored, cf. the distinction in Levinson (2000: 21–27) between

sentence meaning, utterance type meaning, and speaker meaning, or Carston's three-level model of utterance interpretation. However, there is considerable terminological confusion here, as the diagram in Levinson (2000: 195) impressively shows; confusion that has to do with the still unsolved problem of finding demarcation lines or fixing the interfaces between 'what is said' and 'what is meant.' Further discussion of the question of level architecture can be found in Récanati (2004).

Obviously, the second tendency is connected with the widely accepted view that some sort of underdeterminacy thesis is correct, and that there are presemantic pragmatic processes that are input for model-theoretic interpretation (cf. Levinson, 2000: 188), or are necessary to fix full propositional representations (cf. Carston, 2002).

As has become clear, there are still many problems to solve: the status of the maxims of Relevance and Manner, the distinction between GCI and PCI, the status of conventional implicatures, the interaction of implicatures with illocutions and sentence types, to name only a few. Besides, the role that implicatures play in many areas, such as those of language acquisition and language change, awaits much further research.

See also: Cooperative Principle; Irony; Maxims and Flouting; Meaning: Procedural and Conceptual; Relevance Theory; Semantics–Pragmatics Boundary; Speech Acts.

Bibliography

- Atlas J D (2005). *Logic, meaning, and conversation: semantical underdeterminacy, implicature, and their interface*. Oxford: Oxford University Press.
- Autenrieth T (1997). 'Tautologien sind Tautologien.' In Rolf E (ed.) *Pragmatik. Implikaturen und Sprechakte*. Opladen: Westdeutscher Verlag. 12–32.
- Bach K (1994). 'Semantic slack: what is said and more.' In Tsohatzidis S L (ed.) *Foundations of speech act theory*. London/New York: Routledge. 267–291.
- Bach K (1999). 'The myth of conventional implicature.' *Linguistics and Philosophy* 22, 327–366.
- Bach K (2001). 'You don't say?' *Synthese* 128, 15–44.
- Bach K & Harnish R M (1979). *Linguistic communication and speech acts*. Cambridge, MA: The MIT Press.
- Blakemore D (2002). *Relevance and linguistic meaning: the semantics and pragmatics of discourse connectives*. Cambridge: Cambridge University Press.
- Blutner R (2004). 'Pragmatics and the lexicon.' In Horn L R & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 488–514.
- Brown P & Levinson S C (1987). *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
- Carston R (1988). 'Implicature, explicature and truth-theoretic semantics.' In Kempson R (ed.) *Mental representations: the interface between language and reality*. Cambridge: Cambridge University Press. 155–181.
- Carston R (1996). 'Metalinguistic negation and echoic use.' *Journal of Pragmatics* 25, 309–330.
- Carston R (2002). *Thoughts and utterances: the pragmatics of explicit communication*. Oxford: Blackwell.
- Chapman S (1996). 'Some observations on metalinguistic negation.' *Journal of Linguistics* 32, 387–402.
- Clark B (1993). 'Relevance and "Pseudo-Imperatives".' *Linguistics and Philosophy* 16, 79–121.
- Clark E V (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Clark E V (2004). 'Pragmatics and language acquisition.' In Horn L R & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 562–577.
- Dascal M (1994). 'Speech Act Theory and Gricean pragmatics. Some differences of detail that make a difference.' In Tsohatzidis S L (ed.) *Foundations of speech act theory*. London/New York: Routledge. 323–334.
- Davies E E (1986). *The English imperative*. London: Croom Helm.
- Davis W A (1998). *Implicature: intention, convention and principle in the failure of Gricean theory*. Cambridge: Cambridge University Press.
- d'Avis F (2001). *Über 'w-Exklamativsätze' im Deutschen*. Tübingen: Niemeyer.
- Fretheim T (1992). 'The effect of intonation on a type of scalar implicature.' *Journal of Pragmatics* 18, 1–30.
- Gazdar G (1979). *Pragmatics: implicature, presupposition and logical form*. New York: Academic Press.
- Gibbs R W Jr (2002). 'A new look at literal meaning in understanding what is said and implicated.' *Journal of Pragmatics* 34, 457–486.
- Giora R (2003). *On our mind: salience, context, and figurative language*. Oxford: Oxford University Press.
- Grice P (1989). 'Logic and conversation.' In Grice P (ed.) *Studies in the way of words*. Cambridge, MA: Harvard University Press. 22–40.
- Harnish R M (1994). 'Mood, meaning and speech acts.' In Tsohatzidis S L (ed.) *Foundations of speech act theory*. London/New York: Routledge. 407–459.
- Hawkins J A (1991). 'On (in)definite articles: implicatures and (un)grammaticality prediction.' *Journal of Linguistics* 27, 405–442.
- Hirschberg J (1991). *A theory of scalar implicature*. New York: Garland.
- Horn L R (1984). 'Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature.' In Schiffrin D (ed.) *Meaning, form, and use in context: linguistic applications*. Washington, DC: Georgetown University Press. 11–42.
- Horn L R (1989). *A natural history of negation*. Chicago/London: The University of Chicago Press.
- Horn L R (2004). 'Implicature.' In Horn L R & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 3–28.
- Iwata S (1998). 'Some extensions of the echoic analysis of metalinguistic negation.' *Lingua* 105, 49–65.

- König E (1991). *The meaning of focus particles: a comparative perspective*. London: Routledge.
- Krifka M (2002). 'Be brief and vague! And how bidirectional optimality theory allows for verbosity and precision.' In Restle D & Zaefferer D (eds.) *Sounds and systems: studies in structure and change. A Festschrift for Theo Vennemann*. Berlin: de Gruyter. 439–458.
- Leech G N (1983). *Principles of pragmatics*. London/New York: Longman.
- Levinson S C (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson S C (1987). 'Minimization and conversational inference.' In Verschueren J & Bertucelli-Papi M (eds.) *The pragmatic perspective*. Amsterdam: Benjamins. 61–129.
- Levinson S C (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: The MIT Press.
- Matsumoto Y (1995). 'The conversational condition on Horn scales.' *Linguistics and Philosophy* 18, 21–60.
- Meibauer J (1986). *Rhetorische Fragen*. Tübingen: Niemeyer.
- Meibauer J (1990). 'Sentence mood, lexical categorial filling, and non-propositional *nicht* in German.' *Linguistische Berichte* 130, 441–465.
- Meibauer J (1997). 'Modulare Pragmatik und die Maximen der Modalität.' In Rolf E (ed.) *Pragmatik: Implikaturen und Sprechakte*. Opladen: Westdeutscher Verlag. 226–256.
- Meibauer J (2005). 'Lying and falsely implicating.' *Journal of Pragmatics* 38(12).
- Noveck I A (2001). 'When children are more logical than adults: experimental investigations of scalar implicature.' *Cognition* 78, 165–188.
- Papafragou A & Musolino J (2003). 'Scalar implicatures: experiments at the semantics-pragmatics interface.' *Cognition* 86, 253–282.
- Portner P & Zanuttini R (2000). 'The force of negation in wh exclamatives and interrogatives.' In Horn L R & Kato Y (eds.) *Negation and polarity: syntactic and semantic perspectives*. Oxford: Oxford University Press. 193–231.
- Posner R (1980). 'Semantics and pragmatics of sentence connectives in natural language.' In Searle J R, Kiefer F & Bierwisch M (eds.) *Speech act theory and pragmatics*. Dordrecht: Reidel. 168–203.
- Potts C (2005). *The logic of conventional implicatures*. Oxford: Oxford University Press.
- Récenati F (1995). 'The alleged priority of literal interpretation.' *Cognitive Science* 19, 207–232.
- Récenati F (2004). *Literal meaning*. Cambridge: Cambridge University Press.
- Reis M (1991). 'Echo-w-Sätze und Echo-w-Fragen.' In Reis M & Rosengren I (eds.) *Fragesätze und Fragen*. Tübingen: Niemeyer. 49–76.
- Reis M (1999). 'On sentence types in German: an enquiry into the relationship between grammar and pragmatics.' *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* 4, 195–236.
- Rolf E (1994). *Sagen und Meinen. Paul Grices Theorie der Konversations-Implikaturen*. Opladen: Westdeutscher Verlag.
- Romero M & Han C (2004). 'On negative yes/no questions.' *Linguistics and Philosophy* 27, 609–658.
- Rooy R van (2004). 'Signalling games select Horn strategies.' *Linguistics and Philosophy* 27, 493–527.
- Sadock J M (1978). 'On testing for conversational implicatures.' In Cole P (ed.) *Syntax and semantics 9: Pragmatics*. New York: Academic Press. 281–298.
- Sadock J M (2004). 'Speech acts.' In Horn L R & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 53–73.
- Sadock J M & Zwicky A M (1985). 'Speech act distinctions in syntax.' In Shopen T (ed.) *Language typology and syntactic description I: Clause structure*. Cambridge: Cambridge University Press. 155–196.
- Sauerland U (2004). 'Scalar implicatures in complex sentences.' *Linguistics and Philosophy* 27, 367–391.
- Saul J M (2002). 'What is said and psychological reality: Grice's project and relevance theorists' criticisms.' *Linguistics and Philosophy* 25, 347–372.
- Searle J R (1975). 'Indirect speech acts.' In Cole P & Morgan J (eds.) *Syntax and semantics 3: Speech acts*. New York: Academic Press. 59–82.
- Searle J R (1980). 'The background of meaning.' In Searle J, Kiefer F & Bierwisch M (eds.) *Speech act theory and pragmatics*. Dordrecht: Reidel. 221–232.
- Shapiro A M & Murphy G L (1993). 'Can you answer a question for me? Processing indirect speech acts.' *Journal of Memory and Language* 32, 211–229.
- Sperber D & Wilson D (1981). 'Irony and the Use-Mention-Distinction.' In Cole P (ed.) *Radical pragmatics*. New York: Academic Press. 295–318.
- Sperber D & Wilson D (1995). *Relevance. Communication and cognition* (2nd edn.). Oxford: Blackwell.
- Sperber D & Wilson D (1998). 'Irony and relevance: a reply to Seto, Hamamoto and Yamanashi.' In Carston R & Uchida S (eds.) *Relevance theory: applications and implications*. Amsterdam: Benjamins. 283–293.
- Traugott E C (2004). 'A critique of Levinson's view of Q- and M-inferences in historical pragmatics.' *Journal of Historical Pragmatics* 5, 1–25.
- Wilson D & Sperber D (1992). 'On verbal irony.' *Lingua* 87, 53–76.
- Wilson D & Sperber D (2002). 'Truthfulness and relevance.' *Mind* 111, 583–632.
- Wilson D & Sperber D (2004). 'Relevance theory.' In Horn L R & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 607–632.
- Winner E (1988). *The point of words: children's understanding of metaphor and irony*. Cambridge, MA: Harvard University Press.
- Zanuttini R & Portner P (2003). 'Exclamative clauses: at the syntax-semantics-interface.' *Language* 79, 39–81.

Indeterminacy, Semantic

M Hymers, Dalhousie University, Halifax, Nova Scotia, Canada

© 2006 Elsevier Ltd. All rights reserved.

Contemporary philosophy has presented two significant challenges to the determinacy of meaning, one contentiously associated with Ludwig Wittgenstein through the work of Saul Kripke (see **Rules and Rule-Following**) and the other, the topic of this entry, traceable to the behaviorism of W. V. Quine by way of his thought experiment concerning radical translation.

Quine purports to offer two distinct arguments for semantic indeterminacy, which he sees as a consequence of the thesis of the *indeterminacy of translation*: “manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another” (1960: 27). This thesis has semantic implications for Quine, because as a behaviorist he holds that meaning is nothing above and beyond what is preserved in translation. This article will discuss both arguments for this thesis and reactions to them in separate sections below, focusing largely on the so-called argument ‘from below’ (1970: 183).

The Argument from Below

The radical translator, says Quine, is someone who has at her disposal only the behavior of the speaker of the alien tongue she is trying to translate and the observable objects and events that grace the locale she shares with the speaker. More carefully, Quine focuses not on those objects and events themselves, but on the stimulations they produce at the speaker’s nerve endings (see **Radical Interpretation, Translation and Interpretationalism**) and on the speaker’s observed disposition to assent to or dissent from sentences that the translator puts to her as queries under given *stimulus-conditions*. If she can formulate correct hypotheses about the speaker’s terms for affirmation and dissent, then this evidential constraint entails that the field linguist can translate *observation-sentences*, which are among the class of sentences to which assent is cued directly to particular occasions and to which, moreover, assent varies little or not at all with changes in the background information available to the speaker (1960: 42). The thought here is that reports of observations are reports of things observed on particular occasions (unlike talk of biological speciation or social justice), and those

things are importantly independent of the speaker’s background beliefs and mental states (unlike hallucinations, for example). Each such sentence has what Quine calls a *stimulus-meaning* – the ordered pair of its positive stimulus-conditions (those prompting assent to the sentence) and its negative stimulus-conditions (those prompting dissent from the sentence). Logical connectives such as ‘not,’ ‘and,’ ‘or,’ and ‘all’ can also be translated, because “one’s interlocutor’s silliness, beyond a certain point, is less likely than bad translation” (59). This ‘principle of charity’ (59n) has it that, for example, a speaker will not typically affirm both a sentence and its negation. We can, as well, identify but not necessarily translate sentences that are *stimulus synonymous* (those provoking assent and dissent respectively under exactly the same stimulus conditions) and sentences that are *stimulus analytic* – those to which the speaker will assent, if not to nothing, on every stimulus occasion (54–55).

Quine begins with sentences because a speaker assents to (or dissents from) utterances of sentences, not to (or from) isolated subsentential terms. But a thorough ‘translation manual’ should tell us how to match up individual words or other subsentential expressions (1960: 69). The radical translator may thus formulate a set of *analytical hypotheses* (68), which break down sentences into their component parts, words, which can then be found in other sentences. A survey of their occurrences in different sentences makes it possible to pair those words with words of the translator’s language, eventually yielding a comprehensive translation manual, whose adequacy can be further tested by recombining the isolated words into new sentences and observing the speaker’s pattern of assent and dissent to them. However, Quine contends, there will always be more than one overall system of analytical hypotheses that will yield workable translations, because local incompatibilities in competing hypotheses can always be compensated for by making changes elsewhere in the system. Thus, to take a trivial case, in the French ‘*ne...rien*,’ ‘*rien*’ can be translated either as ‘anything’ or as ‘nothing,’ says Quine, “by compensatorily taking ‘*ne*’ as negative or as vacuous” (1969a: 37).

Worse still, even if one’s analytical hypotheses isolate a particular word, the reference of that word may itself remain ‘behaviorally inscrutable’ (1969a: 35), for there is no empirical test that will decide between translations of the word ‘*gavagai*’ as ‘rabbit’ and as ‘undetached rabbit parts,’ given that apparent evidence against one translation can again be compensated for by making adjustments elsewhere in

the system (see **Holism, Semantic and Epistemic**). Inquiries whether this *gavagai* (rabbit) is the same as that *gavagai* (rabbit) may as easily be taken as inquiries whether this *gavagai* (rabbit part) belongs with that *gavagai* (rabbit part) (33). Beyond the level of observation-sentences and logical connectives, Quine contends, translation is radically indeterminate.

Quine assumes that overt bodily behavior is the only relevant evidence for the field linguist's efforts (1990: 37–38; see **Behaviorism: Varieties**), and he allows that the radical translator might have access to the totality of such evidence – that competing translation manuals might “accord perfectly . . . with all dispositions to behavior on the part of all the speakers concerned” (1969a: 29). Therefore, his conclusion is not merely that I cannot *know* what the unique, correct translation of the speaker's utterances is, but that there is *no fact* about what their unique, correct translation is, because all the relevant evidence leaves the question unanswered. Moreover, Quine holds that the same considerations apply to understanding a speaker of my own language – even when that speaker is myself. “On deeper reflection, radical translation begins at home” (46). This is because first as a learner of my own language and then as an interpreter of my colinguists, I have access to no more evidence than the radical translator does. It follows that meaning itself is radically indeterminate.

Reactions to the Argument from Below

Critics have complained that behaviorism is false and, so, that Quine's argument fails to establish its conclusion (Chomsky, 1969); and that Quine, in spite of his avowed behaviorism (Quine, 1990: 37), helps himself to nonbehaviorist data about the speaker's *nonobservational* terms for assent and dissent (Hockney, 1975: 421; Glock, 2003: 178) and about the communicative exchange between speaker and radical translator (Glock, 2003: 175–182). Assent and dissent cannot simply be read from the speaker's behavior, and they cannot intelligibly be assigned stimulus-meanings. The only way that the field linguist can identify terms of assent and dissent is by assuming that the speaker *wants* to be interpreted and understood, that the speaker *understands* that she is being interpreted, that the speaker is familiar with performing the *speech acts* of *answering* a question and *correcting* the linguist's proffered usage, and so on. These assumptions are all ruled out by Quine's behaviorism.

Critics have also doubted whether there really could be entire alternative translation manuals of the sort Quine envisions. The contention, very roughly, is that the compensatory changes that Quine imagines making to other terms in one's translation

manual in order to preserve one's favored translation of some given term would quickly come into conflict with the behavior of the speaker (Evans, 1975: 345–346). If I translate *gavagai* as ‘undetached rabbit part’ instead of as ‘rabbit,’ for example, I may have to translate the speaker as allowing that an undetached rabbit part is white if and only if the entire rabbit is white (Miller, 1998: 139). But this has consequences for how I should translate the speaker's word for ‘white’ in other contexts, and I should not be surprised to discover that I want to translate her as affirming that snow is white, even when visible patches of it have been visited by local sled dogs.

The Argument from Above

Quine might try to meet some of these objections by appealing to the argument ‘from above’ (1970: 183). Physical theory, says Quine, is underdetermined by the available evidence. More than one set of physical hypotheses could be compatible with the totality of empirical evidence. However, even if we select a particular physical theory on pragmatic grounds, Quine contends, our linguistic theory will remain underdetermined relative to our physical theory – “even if we ‘fix’ the physics” (Putnam, 1978: 53), nonobservational meaning will remain underdetermined. Linguistic hypotheses are underdetermined by a set of *more basic* ‘facts’ that are themselves already underdetermined. It is this double underdetermination that distinguishes semantic indeterminacy from normal empirical underdetermination.

Reactions to the Argument from Above

It is not clear, however, that this second argument operates independently of the argument from below, for the only plausible reason for thinking that the linguistic data are doubly underdetermined with respect to physical theory seems to be that mere physical behavior and the observation-sentences it undergirds are more evidentially basic – the only evidence we have to go on – and rival systems of analytical hypotheses are compatible with that evidence (Kirk, 1986: 140–146; Miller, 1998: 147). Moreover, as Chomsky complains, even if linguistics remains underdetermined once the physics has been fixed, it is equally true that the physics remains underdetermined once the linguistics has been fixed (Chomsky, 1980: 20; Rorty, 1972: 451–453). Only if we grant physics a special privilege could there be a problem for the determinacy of meaning.

Quine's work has, nonetheless, been very influential in analytical philosophy of language, leading especially to the varieties of *interpretationalism*

defended by Donald Davidson and Daniel Dennett (see **Radical Interpretation, Translation and Interpretationalism**).

See also: Behaviorism: Varieties; Compositionality: Philosophical Aspects; Holism, Semantic and Epistemic; Radical Interpretation, Translation and Interpretationalism; Rules and Rule-Following.

Bibliography

- Chomsky N (1969). 'Quine's empirical assumptions.' In Davidson D & Hintikka J (eds.). 53–68.
- Chomsky N (1980). *Rules and representations*. New York: Columbia University Press. Chapter 1.
- Davidson D & Hintikka J (eds.) (1969). *Words and objections: essays on the work of W. V. Quine*. Dordrecht: D. Reidel.
- Evans G (1975). 'Identity and Predication.' *The Journal of Philosophy* 72, 343–363.
- Gibson R F (1982). *The philosophy of W. V. Quine: an expository essay*. Tampa: University of South Florida, 63–95, 176–205.
- Glock H-J (2003). *Quine and Davidson on language, thought and reality*. Cambridge: Cambridge University Press. Chapters 6–7.
- Hockney D (1975). 'The bifurcation of scientific theories and indeterminacy of translation.' *Philosophy of Science* 42, 411–427.
- Hookway C (1988). *Quine: language, experience and reality*. Stanford: Stanford University Press. Chapters 8–10.
- Kirk R (1986). *Translation determined*. Oxford: Clarendon Press.
- Martin R M (1987). *The meaning of language*. Cambridge, MA: MIT Press. Chapter 6.
- Miller A (1998). *Philosophy of language*. Montreal: McGill-Queen's University Press. Chapter 4.
- Putnam H (1978). *Meaning and the moral sciences*. London: Routledge & Kegan Paul. Lecture IV.
- Quine W V (1960). *Word and object*. Cambridge, MA: MIT Press. Chapters I–II.
- Quine W V (1969a). *Ontological relativity and other essays*. New York: Columbia University Press.
- Quine W V (1969b). 'Reply to Chomsky.' In Davidson D & Hintikka J (eds.). 302–311.
- Quine W V (1970). 'On the reasons for the indeterminacy of translation.' *The Journal of Philosophy* 67, 178–183.
- Quine W V (1990). *Pursuit of truth*. Cambridge, MA: Harvard University Press. Chapter III.
- Rorty R (1972). 'Indeterminacy of translation and of truth.' *Synthese* 23, 443–462.

Indexicality: Philosophical Aspects

E Corazza, University of Nottingham,
Nottingham, UK

© 2006 Elsevier Ltd. All rights reserved.

In most of our linguistic interchanges and thinking episodes, we rely on context to select items of discourse and items of thought. One often succeeds in talking and thinking about something because one is situated in a given context. In natural language we have tools whose specific function is to exploit the context of use in order to select an item in one's surroundings. If one says, 'It is raining here' while in London, one refers to London because one's utterance occurs in London. Were one to utter the same sentence in Paris, one would be referring to Paris. We can use the very same words and yet refer to very different items. When you use 'I', for instance, you refer to yourself, whereas when I use it, I refer to myself. We use the very same linguistic expression with the same conventional meaning. It is a matter of who uses it that determines *who* the referent is. Moreover, when Ivan, pointing to Jane, says 'she' or 'you,' he refers to Jane; Jane, however, cannot refer to herself using 'she'

or 'you' (unless she is addressing an image of herself). If we change the context – the speaker, the time, the place – in which these expressions occur, we may end up with a different referent.

Among the expressions that may switch reference with a change in context, we have personal pronouns ('my', 'you', 'she', 'his', 'we'...), demonstrative pronouns ('this', 'that'), complex demonstratives ('this pencil', 'that brunette in the corner'...), adverbs ('today', 'yesterday', 'now', 'here'...), adjectives ('actual', 'present'), and possessive adjectives ('my pencil', 'their car'...).

These expressions have been termed, following Peirce, *indexicals*. Indexicals constitute the paradigm of context-sensitive expressions, i.e., those expressions that rely on the context of use to select an object of discourse. Reichenbach (Reichenbach, 1947) claimed that indexicals are *token reflexive*, for they can be defined in terms of the locution 'this token', where the latter (reflexively) self-refers to the very token used. So, 'I' can be defined in terms of 'the person who utters this token', 'now' in terms of 'the time at which this token is uttered', 'this pen' in terms of 'the pen indicated by a gesture

accompanying this token', etc. The reference of an indexical expression depends on its particular linguistic meaning: 'the utterer of this token' is the linguistic meaning (the *character* (Kaplan) or *role* (Perry)) of 'I', while 'the day in which this token is uttered' is the linguistic meaning of 'today', and so on. The meaning of an indexical can be viewed as a rule which one needs to master to use an indexical correctly. An indexical's linguistic meaning can be conceived as a function taking as its argument the context and giving as its value the referent/content (this is Kaplan's famous content/character distinction).

It is often the case, however, that the linguistic meaning of expressions such as 'this', 'that', 'she', etc., together with context, is not enough to select a referent. These expressions are often accompanied by a pointing gesture or demonstration, and the referent will be what the demonstration demonstrates. Kaplan (1977) distinguishes between *pure indexicals* ('I', 'now', 'today', ...) and *demonstratives* ('this', 'she', ...). The former, unlike the latter, do not need a demonstration – or *directing intention*, Kaplan (1989) – to secure the reference.

In their paradigmatic use, pure indexicals differ from demonstratives insofar as the latter, unlike the former, are perception based. When one says 'I', 'today', etc., one does not have to perceive herself or the relevant day to competently use and understand these expressions. To competently use and understand 'this', 'she', etc., one ought to perceive the referent or demonstratum. For this reason, when a pure indexical is involved, the context of reference fixing and the context of utterance cannot diverge: the reference of a pure indexical, unlike the reference of a demonstrative, cannot be fixed by a past perception.

Moreover, a demonstrative, unlike a pure indexical, can be a vacuous term. 'Today', 'I', etc., never miss the referent. Even if I do not know whether today is Monday or Tuesday and I am an amnesiac, when I say 'Today I am tired,' I refer to the relevant day and to myself. By contrast, if one says 'She is funny' while hallucinating, or 'This car is green' while pointing to a man, 'she' and 'this car' are vacuous.

In addition, pure indexicals cannot be coupled with sortal predicates, while 'this' and 'that' often are used to form complex demonstratives such as 'this *book*', 'that *water*'. Sortal predicates can be considered to be *universe narrowers* which, coupled with other contextual clues, help us to fix a reference. If one says 'This liquid is green' while pointing to a bottle, the sortal 'liquid' helps us to fix the liquid and not the bottle as the referent. Moreover, personal pronouns which work like demonstratives (e.g., 'she', 'he', 'we',) have a built-in or hidden sortal. 'She', unlike

'he', refers to a female, while 'we' usually refers to a plurality of people, of whom one will be the speaker.

Indexicals are generally conceived of as singular terms that contribute a referent to what is said. According to the *direct reference* view (from Kaplan and Perry), utterances containing indexicals express singular propositions, i.e., propositions whose constituents are the referents of the indexicals. As such, indexicals are usually characterized as expressions whose interpretation requires the identification of some element of the utterance context, as stipulated by their linguistic meaning. Thus, an utterance of 'I am tired' expresses a proposition containing the referent of the first person pronoun, and one understands it insofar as one knows to whom the term 'I' refers in the context in which it is uttered. The linguistic meaning governing the use of the indexical – such as 'the agent of the utterance' *qua* meaning of 'I', 'the day of the utterance' *qua* meaning of 'today' – does *not* feature as a constituent of the proposition expressed.

If indexical expressions are characterized as singular terms contributing their referents into what is said (i.e., the proposition expressed), adjectives such as 'local', 'distant', 'actual' – not to mention count nouns like '(a) foreigner', '(an) enemy', '(an) outsider', '(a) colleague' – would not fall into the same category, for they do not contribute a referent to the proposition expressed. Yet they are, plausibly, context-sensitive expressions. 'Local', 'foreign', and 'native' in 'A *local* bar is promoting *foreign* wine' and 'A *native* speaker should correct your essay' do not contribute a specific individual or individuals to the proposition expressed. Hence, they are not singular terms. It should be evident that context-sensitivity does not merely concern singular terms. It is worth distinguishing between indexicals *qua* singular terms, contributing their referents to the proposition expressed, and *contextuals qua* expressions which, though context-sensitive, are not singular terms. Adjectives such as 'tall', 'big', 'small', 'old', etc., also are context-sensitive, insofar as one is only tall/small/big/old ... relative to a comparison class. Jon may be too short to play tennis and yet too tall to be a jockey, while Jane may be too old to join the army and too young to take early retirement. But see Cappelen and Lepore (2004) and Borg (2004) for the view that words such as 'tall', 'foreigner', 'old', and the like are not genuinely context sensitive.

Proper names, like indexicals, also contribute individuals into the proposition expressed. As such they are singular terms, too; yet they are not indexicals (but see Recanati, 1993 for a different view). Nouns such as 'Monday', 'February', and the like also seem to contribute specific individuals in the proposition

expressed. They are best viewed in the same light as count nouns, i.e., as nouns such as 'lemon', 'frog', and 'table' (see Corazza, 2004). As such, they can be used to build singular terms. This happens when they are coupled with an indexical expression such as 'this', 'next', 'last' and they contribute to complex demonstratives of the form 'next week', 'last Saturday', 'next Christmas'. This peculiarity parallels the way count nouns can participate in building complex demonstratives such as 'these lemons', 'that table', 'this pen'. (King, however, defends the view that complex demonstratives are quantified terms).

One of the major features of indexicals differentiating them from other referential expressions (e.g., proper names: 'Plato', 'Paris'; mass terms: 'silver', 'water', terms for species: 'frog', 'raspberry', and so on) is that they are usually used to make reference *in praesentia*. That is, use of an indexical exploits the presence of the referent. Usually in a communicative episode involving an indexical, the referent is in the perceptual field of the speaker and contextual clues are used to raise the referent to salience (see Smith, 1989; Sidelle, 1991; and Predelli, 1998 for a discussion of indexicals used to refer to objects not present, e.g., answering machines, post-it notes, etc.)

When indexicals are not used to make reference *in praesentia* they exploit a previously fixed reference. 'That boy' in 'That boy we encountered yesterday was in trouble with the police' does not refer to someone present. In cases like this, the indexical makes reference *in absentia*. One can thus distinguish between the context of utterance and the context of reference fixing. In our example, the speaker and the hearer appeal to a past context to fix the reference. The gap between the two contexts would be bridged by memory. Another way to handle examples like this would be to argue that, in such cases, the indexical expression works like an anaphoric pronoun linked to a tacit initiator. In the sentence 'In 1834 Jane visited her parents, *now* two old, sick people,' 'now' does not refer to the time of the utterance. It refers to 1834. It does so because it is anaphorically linked to '1834', and, as such, it inherits its reference from it. A similar story could be told about 'that boy': it inherits its reference from a tacit initiator, i.e., an unpronounced NP which is nonetheless presupposed in the discourse situation. To stress this interpretation, consider the following exchange: Jane: 'It is raining'; Jon: 'Then I won't be there before tomorrow.' In saying 'It is raining,' Jane tacitly refers to the location she is in, say London. With 'there', Jon refers to the very same location and we can claim that he does so because 'there' works in an anaphoric way, inheriting its value from the tacit reference made by Jane.

Furthermore, indexicals differ from other referential expressions insofar as (in their paradigmatic use, at least) they cannot be deferential. While one often relies on the so-called division of linguistic labor when using non-indexical expressions (e.g., proper names or mass terms), one cannot depend on the same phenomenon when using an indexical. One can, for instance, competently use 'Feynman' or 'elm' even if one does not know who Feynman is and even if one is unable to tell an elm from a pine. Indeed, a blind person can utter 'that vase' when she has been told that there is a vase in front of her. In these uses the reference is fixed by someone else (it is deferentially fixed). However, these are far from being the paradigmatic uses of an indexical such as 'that/this'. In their paradigmatic uses, they refer to something the user is perceptually aware of. This difference between indexicals and other terms parallels the fact that when using proper names, mass terms, and the like, context is in play before the name is used. As Perry suggests, we often use context to disambiguate a mark or noise (e.g., 'bank', or 'Socrates' used either as a tag for the philosopher or for the Brazilian football player). These are pre-semantic uses of context. With indexicals, though, context is used semantically. It remains relevant after the language, words, and meaning all are known; the meaning directs us to certain aspects of context. This distinction reflects the fact that proper names, mass terms, etc., unlike indexicals, contribute to building context-free (eternal) sentences, that is, sentences that are true or false independently of the context in which they are used.

To sum up, philosophers have made several key claims about indexicals. They are tools whose function is to exploit context, and their hallmarks include not having a fixed referent, not being easily deployed *in absentia* of the thing referred to, not being used deferentially, and having context play (not just a pre-semantic role, i.e., determining which word has been used, but also) a semantic role. Philosophers have found that indexicals come in at least three varieties: pure indexicals ('I', 'now'), demonstratives ('this', 'she'), and contextuais ('foreign', 'local'). Key differences between the first and second variety are that, in contrast to pure indexicals, demonstratives are more perception-based, they may be vacuous, they can be combined with sortals, and directing intentions play a quite central role in their use. In addition to attracting the attention of philosophers, indexicals have also captured the interest of those working within the boundaries of cognitive science for several reasons (see, for instance, Pylyshyn, 2003 on how indexicality is relevant to the study of vision): they play crucial

roles when dealing with such puzzling notions as the nature of the self (see for instance the importance of 'I' in Descartes' cogito argument), the nature of perception, the nature of time, psychological pathologies, social interaction, and psychological development (see Corazza, 2004).

See also: Pragmatic Determinants of What Is Said; Semantic Value.

Bibliography

- Austin D (1990). *What's the meaning of "this"?* Ithaca: Cornell University Press.
- Bach K (1987). *Thought and reference*. Oxford: Clarendon Press.
- Biro J (1982). 'Intention, Demonstration, and Reference.' *Philosophy and Phenomenological Research* 43, 35–41.
- Borg E (2004). *Minimal semantics*. Oxford: Oxford University Press.
- Cappelen H & Lepore E (2004). *Insensitive semantics*. Oxford: Basil Blackwell.
- Castañeda H-N (1989). *Thinking, language, and experience*. Minneapolis: University of Minnesota Press.
- Chisholm R (1981). *The first person*. Minneapolis: University of Minnesota Press.
- Corazza E (2003). 'Complex Demonstratives *qua* Singular Terms.' *Erkenntnis* 59, 263–283.
- Corazza E (2004). *Reflecting the mind: Indexicality and quasi-indexicality*. Oxford: Oxford University Press.
- Evans G (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Frege G (1918). 'Thoughts.' In Salmon N & Soames S (eds.) (1988). *Propositions and attitudes*. Oxford: Oxford University Press. 33–55.
- Kaplan D (1977). 'Demonstratives.' In Almog J *et al.* (eds.) (1989). *Themes from Kaplan*. Oxford: Oxford University Press. 481–563.

- Kaplan D (1989). 'Afterthoughts.' In Almog J *et al.* (eds.) (1989). *Themes from Kaplan*. Oxford: Oxford University Press. 565–614.
- King J (2001). *Complex demonstratives: A quantificational account*. Cambridge: MIT Press.
- Lewis D (1979). 'Attitudes *de dicto* and *de se*.' *The Philosophical Review* 88, 513–543. Reprinted in Lewis D (1979). *Philosophical Papers*, vol. 1. Oxford: Oxford University Press.
- Mellor D H (1989). 'I and Now.' *Proceeding of the Aristotelian Society* 89, 79–84. Reprinted in Mellor D H (1989). *Matters of metaphysics*. Cambridge: Cambridge University Press. 79–84.
- Numberg G (1993). 'Indexicality and Deixis.' *Linguistics and Philosophy* 16, 1–43.
- Perry J (2000). *The problem of the essential indexical and other essays*. Stanford: CSLI Publications.
- Perry J (2001). *Reference and reflexivity*. Stanford: CSLI Publications.
- Predelli S (1998). 'Utterance, Interpretation, and the Logic of Indexicals.' *Mind and Language* 13, 400–414.
- Pylyshyn Z (2003). *Seeing and visualizing*. Cambridge: MIT Press.
- Recanati F (1993). *Direct reference*. London: Blackwell.
- Reichenbach H (1947). *Elements of symbolic logics*. New York: Free Press.
- Sidelle A (1991). 'The answering machine paradox.' *Canadian Journal of Philosophy* 81, 525–539.
- Smith Q (1989). 'The multiple uses of indexicals.' *Synthese* 78, 67–91.
- Vallée R (1996). 'Who are we?' *Canadian Journal of Philosophy* 26, 211–230.
- Yourgrau P (ed.) (1990). *Demonstratives*. Oxford: Oxford University Press.
- Wettstein H (1991). *Has semantics rested on a mistake? and other essays*. Stanford: Stanford University Press.

Innate Ideas

E Margolis, Rice University, Houston, TX, USA

S Laurence, University of Sheffield, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

The question of whether there are any innate ideas or concepts goes back to ancient times and is at the center of the dispute between empiricists and nativists. Empiricists maintain that the mind has little innate structure and that all of our concepts are based on perception. In contrast, nativists maintain that the mind has a great deal of innate structure, including a stock of innate concepts of considerable abstraction (e.g., CAUSE, AGENT, NUMBER).

The traditional dispute between empiricists and nativists, especially in the 17th and 18th centuries, often confounded psychological issues with epistemological ones. While the former address how the mind is organized and how it develops, the latter address the grounds for justifying our beliefs. Contemporary theorists view the psychological issues as being largely independent of the epistemological issues. Contemporary theorists also draw upon empirical studies of the mind to support their claims. The cognitive sciences – from linguistics to developmental psychology to cognitive anthropology – all bear on the innateness controversy. At the same time, questions about innate concepts continue to be hotly debated

among philosophers. One issue concerns what it means to say that a concept is innate and whether the notion of innateness has explanatory value. A second and related issue concerns which concepts (if any) are innate and how arguments for innate concepts are to be evaluated.

What Is Innateness?

Empiricists and nativists disagree about the innate structure of the mind. The disagreement is sometimes mischaracterized by saying that nativism is the view that the mind has an innate structure. This is a mistake, because the point of contention isn't about whether or not something is innate but is about how rich our innate endowment is. Even the most diehard empiricists maintain that the mind has an innate structure that is partly responsible for why humans develop the concepts they do. Even the behaviorist, as W. V. O. Quine notes, "is knowingly and cheerfully up to his neck in innate mechanisms" (1969, pp. 95–96). What distinguishes empiricists from nativists is that empiricists think that there are no innate concepts and that the innate mechanisms that explain how we acquire our concepts are general-purpose mechanisms (for empiricists the same mechanisms, ultimately based on our perceptual systems, are responsible for the development of such diverse concepts as NUMBER, BELIEF, SQUIRREL, and AXE). In contrast, nativists embrace innate concepts and allow that the mind may have considerable innate differentiation.

Some theorists are skeptical of the dispute between empiricists and nativists because they think that little sense can be made of the idea that anything is innate. This poses a potential challenge (to both nativists and empiricists) to clarify the notion of innateness.

One proposal is that innate concepts are ones that are genetically determined. Unfortunately, the notion of genetic determination is itself problematic. Every phenotypic trait is dependent upon both genetic and environmental factors (including the embryonic environment); none can be traced to the genes uniquely. In addition, the appeal to genetic determination relies on a misleading picture that a genotype is a blueprint specifying a detailed sketch of an organism's phenotypic properties. But according to modern genetics, the relationship between genes and phenotypes is far more indirect (Sterelny and Griffiths, 1999; Marcus, 2004).

Another proposal is that innate concepts involve a greater influence of the genes than the environment in the sense that innate concepts are bound to emerge in a wide range of environments (Sober, 1998). A related proposal is that a concept is innate to the extent that

its development is canalized; that is, there is a developmental process that is bound to form the concept regardless of environmental variation (Ariew, 1996). One problem with these sorts of approaches is that some concepts are likely to emerge across a wide range of environments not because they are innate but simply because they are readily *learned* in a wide range of environments. A second problem is that for an account of this kind to work we must be given a principled specification of the relevant environmental variation (Cowie, 1999; Samuels, 2002).

Another approach is to explain innateness in terms of the kind of explanation that accounts for a concept. Samuels (2002) suggests that a psychological structure is innate when it is posited by a correct psychological theory and yet there isn't a psychological account of how it is acquired. For example, a concept is innate if we must turn to a neurobiological explanation of its acquisition. Of course, this leads to the problem of clarifying what counts as a psychological (as opposed to a nonpsychological) explanation, but perhaps this distinction is well enough understood in practice.

Which Concepts are Innate?

Much of the contemporary discussion about which concepts are innate has centered around Jerry Fodor's notorious claim that all **lexical concepts** are innate. Roughly, lexical concepts are ones that are encoded by single words in natural language, as opposed to phrases – for example, the concept DOG is a lexical concept.

Fodor's original presentation of his argument has the feel of a paradox (Fodor, 1975). The argument begins with the distinction between learning and other fortuitous ways of acquiring a concept (e.g., bumping your head in just the right way or undergoing futuristic neurosurgery). Fodor claimed that learning, in contrast with these other ways of acquiring a concept, should be understood as a rational process in which hypotheses are tested. The problem, as Fodor saw it, is that hypothesis testing requires that one already have the concept one is attempting to learn. He points out that in so-called concept learning experiments, subjects are required to sort stimuli in accordance with a novel concept. But if the correct hypothesis is, for example, that the concept in question is the concept of something that is blue or square, then one would have to have prior possession the concept BLUE OR SQUARE – the concept that is supposed to be learned – in order to even be in a position to frame and test the hypothesis. In later work, Fodor refined his argument by allowing that

learning in some sense can take place so long as the acquired concept is complex. Learning is understood to be a process in which a new concept is constructed from previously available constituents in the course of framing and testing a hypothesis (Fodor, 1981). The problem now, however, is that the evidence is supposed to suggest that lexical concepts aren't complex, that they are actually atomic or primitive concepts (*see Concepts*). Fodor concluded that lexical concepts can't be learned, and hence that they must be innate.

The logic of Fodor's argument has been endorsed by such prominent cognitive scientists as Ray Jackendoff and Steven Pinker, who have agreed that concept learning can only be a process where complex concepts are assembled from their constituents (Jackendoff, 1989; Levin and Pinker, 1991). Where they have disagreed with Fodor is about the evidence regarding conceptual structure, maintaining that lexical concepts do in fact have the structure that would allow them to be learned. In contrast, philosophers have often simply dismissed Fodor's argument on the grounds that his strong form of nativism is inconsistent with evolution (Putnam, 1988). They point out that natural selection couldn't have given us an innate conceptual system that incorporates concepts of no utility to our ancestors (e.g., PROTON and SAXOPHONE).

Both of these reactions are unsatisfactory. The appeal to evolution barely addresses Fodor's argument. Simply dismissing his conclusion tells us nothing about where the argument goes wrong and gives us no insight whatsoever into how concepts can be learned. The strategy of localizing the disagreement to one about conceptual structure is certainly better. But it still embraces Fodor's fundamental assumption that there is no way of expanding the combinatorial expressive power of the conceptual system.

Another approach to Fodor's argument embraces the challenge of showing how a primitive concept can be learned (Margolis, 1998; Laurence and Margolis, 2002). The way to do this is to frame the issue of learning side by side with a theory of content that is suitable for primitive concepts, e.g., a causal theory of content. Roughly, on a causal theory of content, a concept's content is fixed by its standing in a nomic-causal relation with the property it expresses. These nomic relations aren't basic laws, so they must be supported by mechanisms that sustain the link between a concept and its corresponding property – sustaining mechanisms. Typical sustaining mechanisms involve inferential dispositions, but because of the assumption of conceptual atomism, different people (and the same person at different times) can have different sustaining mechanisms for the same

concept. What matters for the possession of a primitive concept is the property it is nomically connected to, not the mechanism that establishes the connection. The question of how a concept is acquired then becomes the question of how these sustaining mechanisms are established.

For natural kind concepts, one important type of sustaining mechanism may combine an appreciation of the typical appearance of a kind with a general disposition towards psychological essentialism (the view that instances of a natural kind have a common set of properties that bind them together and that account for the kind's appearance but are themselves not readily observable; see Gelman, 2003). If this is right, then acquiring a natural kind concept can be explained roughly as follows. Certain cues suggest that an object is an instance of natural kind; one is then disposed to keep track of its salient physical characteristics; and this information is organized around a new representation to interact with the general essentialist disposition. The result is a representation that is elicited by things with a similar appearance but that is withheld in those cases where the evidence suggests that the essential properties are missing.

It's natural to describe this whole process as one where a concept is learned, so it illustrates how a primitive concept *can* be learned. However, it doesn't follow that empiricism has been vindicated. Concept learning along these lines may depend upon innate cognitive mechanisms of considerable richness. For example, the model presupposes a disposition towards psychological essentialism and an ability to track natural kinds. These may very well be innate.

Moreover, nativist research has been flourishing in recent years with its focus on areas of cognition that are thought to be subserved by domain-specific systems. These are cases of distinct cognitive mechanisms that are differentiated in their developmental trajectory, that employ specialized computational operations, that draw upon a distinctive body of information, and that may be associated with specialized neural circuits. Some suggestions along these lines are a domain for reasoning about minds, a domain for reasoning about physical objects, a domain for reasoning about biological phenomena, and a domain for reasoning about social exchanges (Hirschfeld and Gelman, 1994; Carruthers *et al.*, 2005). Connected with these domains are strong candidates for innate concepts, including BELIEF, PHYSICAL OBJECT, ANIMAL, and CHEATER.

See also: Causal Theories of Reference and Meaning; Cognitive Science and Philosophy of Language; Concepts; Definitions: Uses and Varieties of; Empiricism;

Externalism about Content; Innate Knowledge; Mentalism; Natural Kind Terms; Representation in Language and Mind.

Bibliography

- Ariew A (1996). 'Innateness and canalization.' *Philosophy of Science* 63 (Proceedings), S19–S27.
- Carruthers P, Laurence S & Stich S (2005). *The innate mind: structure and content*. New York: Oxford University Press.
- Cowie F (1999). *What's within? Nativism reconsidered*. Oxford: Oxford University Press.
- Fodor J A (1975). *The language of thought*. New York: Thomas Y. Crowell.
- Fodor J A (1981). 'The present status of the innateness controversy.' In *Representations: philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press. 257–316.
- Gelman S (2003). *The essential child*. Oxford: Oxford University Press.
- Hirschfeld L & Gelman S (1994). *Mapping the mind*. Cambridge: Cambridge University Press.
- Jackendoff R (1989). 'What is a concept, that a person may grasp it?' *Mind & Language* 4, 68–102.
- Laurence S & Margolis E (2002). 'Radical concept nativism.' *Cognition* 86, 25–55.
- Levin B & Pinker S (1991). 'Introduction.' In Levin B & Pinker S (eds.) *Lexical & conceptual semantics*. Cambridge, MA: Blackwell.
- Marcus G (2004). *The birth of the mind: how a tiny number of genes creates the complexity of the human thought*. New York: Basic Books.
- Margolis E (1998). 'How to acquire a concept.' *Mind & Language* 13, 347–369.
- Putnam H (1988). *Representation and reality*. Cambridge, MA: MIT Press.
- Quine W V O (1969). 'Linguistics and philosophy.' In Hook S (ed.) *Language and philosophy: a symposium*. New York: New York University Press. 95–98.
- Samuels R (2002). 'Nativism and cognitive science.' *Mind & Language* 17, 3, 233–265.
- Sober E (1998). 'Innate knowledge.' In *The Routledge encyclopedia of philosophy* 4, 794–797. London: Routledge.
- Sterelny K & Griffiths P (1999). *Sex and death: an introduction to philosophy of biology*. Chicago: University of Chicago Press.

Innate Knowledge

T Simpson, University of Sheffield, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

Debates over the existence and nature of innate knowledge have a long and complex history, in both philosophy and psychology in general and in language and linguistics in particular (see, e.g., Stich, 1975). As one might expect, such debates have produced many distinct and competing notions of innate knowledge and have also resulted in much disagreement over the existence of either necessary or contingent connections between any particular proposals (see, e.g., Griffiths, 2002). However, there are two senses of innate knowledge that consistently dominate both classical and contemporary work. First, that some set of 'ideas' or 'justified true beliefs' are present at birth, and, second, that there exists some set of domain-specific psychological mechanisms, capacities, faculties, or representations which are in some to-be-specified way biologically or genetically pre-determined.

Defense of innate knowledge in the first sense arguably began with Plato, who claimed that as we can be shown to possess 'genuine knowledge' for which experience is insufficient, then such knowledge must ultimately be something 'recollected' from what

we already know. Philosophers of the Enlightenment also considered in detail the question of what kinds of knowledge may be innate, and in so doing questioned what the very notion of innate might mean. In direct contrast to Plato, John Locke (1690) argued that there are, in fact, "no innate principles in the mind" because, among other things, no useful meaning can be given to the notion of innate. Locke argued that if there were any innate principles, then either these principles must be present in our minds from birth, or there must be some 'mark' by which we can distinguish in adult minds those principles which are innate from those acquired by experience. But, Locke claimed, any supposedly innate principles are clearly not present in the minds of infants and 'idiots,' nor can the required marks be found. Thus, Locke concluded, no useful meaning can be given to the notion of innate in the context of explanations of the origins and development of human knowledge.

Note, however, that Locke's arguments here speak only to an understanding of innate knowledge in the first sense given above. Thus even if such arguments are correct, it remains entirely possible that there exists innate knowledge in the second specified sense. Moreover, there are now sound theoretical and empirical reasons to believe that innate knowledge in this second sense does actually exist (see Carruthers

et al., 2005, for a more detailed discussion of what follows).

Physiological traits such as teeth or pubic hair can be usefully understood as innate in this second sense, and there seems no reason to think that some psychological traits could not be similarly so understood. Furthermore, since the middle of the 20th century, results from research in linguistics and other areas of cognitive science, inspired by the pioneering work of Noam Chomsky (1965, 1967) appear to show that there exist various specific psychological traits which necessarily cannot be acquired via individual experience – as well as other traits which perhaps could be so acquired but which demonstrably are not. Thus there do indeed appear to exist some psychological traits that bear a ‘mark’ that distinguishes them from those that are experientially acquired. So much so, in fact, that for many core cognitive domains (e.g., language, naive physics, and number) virtually all contemporary researchers and theorists accept that there exists some degree of domain-specific innate knowledge – in the second sense – relevant to the domain in question. As Chomsky notes, all contemporary linguists agree that “the question is not whether innate structure is a prerequisite for learning, but rather what it is” (1980: 310), and it is with regard to the specific details of this innate structure that debate between linguistic nativists and non-nativists now occurs.

Evidence for an innate language faculty comes from a wide variety of theoretical and empirical sources, most of which are driven by what have been termed ‘arguments from the poverty of the stimulus’ (see, e.g., Chomsky, 1967, 1981; Laurence and Margolis, 2001; Pinker, 1994). These arguments claim to show, through a variety of different means, that the linguistic competence acquired by normally developing children cannot be – or demonstrably is not being – acquired using predominantly empiricist or general-purpose learning methods and mechanisms. Rather, nativists claim, these arguments show that there must exist a considerable amount of innate knowledge which enables children to bridge the ‘gap’ between the language-related input they receive, and their resultant linguistic abilities. Indeed, when one considers the way in which research in linguistics has proceeded in recent years, it rapidly becomes clear just how large this gap is, and thus just how much innate linguistic knowledge children must actually possess in order to bridge it. Contemporary linguists have discovered a huge number of subtle, but nonetheless fundamental, regularities in our linguistic behavior and have put forward many different theories concerning the structure of language. Moreover, linguists have yet to comprehensively determine

the grammar of even a single human language. This alone suggests that children’s linguistic input is too impoverished to provide them with data sufficient to reach the one correct grammatical hypothesis, which they nonetheless reliably reach by the age of six years old.

In addition, there are theoretical decisions which the developing child needs to make – and reliably makes – but with which linguists have struggled for some considerable time. For example, are rules construction-specific or is sentence structure dictated by the interaction of a number of non-construction-specific rules? Are rules optional or mandatory? Do rules apply in a fixed order, or are they unordered? And so on. There is in fact a theoretically enormous range of grammatical hypotheses that the child could consider during language acquisition, but all normal children nonetheless always arrive at exactly the right hypothesis at approximately the same age. Moreover, all normal children also arrive at exactly the right hypothesis via pretty much the same route. If children were largely empiricist learners, one would expect that each child would try out a huge number of different grammars during language development, and also that the types of mistakes children would make would be highly variable. However, the sorts of errors children make are in fact highly circumscribed (Pinker, 1994; Crain and Pietroski, 2001). This (and much other) evidence therefore overwhelmingly supports the claim that there exists some degree of innate linguistic knowledge. The greater the amount and complexity of such innate knowledge one believes there is, the more nativist one is about our language abilities.

With regard to the actual content of our innate knowledge, many theorists (in linguistics and elsewhere) first draw a distinction between the information that cognition in any given domain requires, and the computational processes which operate on this information in that domain. More specifically, theorists often refer to representational modules and computational modules respectively (see, e.g., Fodor, 1983, 2000; Pinker, 1997; Samuels, 2000). The issue of what makes a cognitive capacity or faculty modular is complex and somewhat controversial (*see Modularity*), but to a first approximation, representational modules are domain-specific bodies of data (where this data is organized and integrated in the right kind of way), and computational modules are domain-specific processing devices. Thus, for instance, “a parser might be conceived of as a computational module that deploys the contents of a [representational] module devoted to linguistic information in order to generate syntactic and semantic representations of physical sentence

forms” (Samuels, 2000: 19). Similar distinctions and consequent models can be generated in various other cognitive domains. Once these distinctions are made, however, it becomes clear that some components of either or both kinds of module could be innate in any given domain, and thus that specifying the details of our innate endowment will require considerable care. In addition, as our innate linguistic endowment will involve at least some extremely specialized representations and computations specific to human grammar, there is no reason to expect there to be any direct isomorphism between the details of our innate linguistic endowment and those of our innate endowment in other cognitive domains. Claims about the details of the innate knowledge appropriate to any given cognitive domain must therefore ultimately be made only in the light of the appropriate domain-specific investigation.

Furthermore, while it is clearly the case that not all language-specific representations or computational processes are innate (i.e., children clearly determine which representations and processes they should actually use on the basis of the input from their local environment), there is also no reason to believe that all the innate components required for language acquisition are language-specific. That is, successful language acquisition may also require the operation of innately specified but more domain-general processes, such as ‘statistical samplers,’ which can or do operate on any auditory or other input. However, the need for such domain-general processes does not detract from the claim that some innate language-specific components also exist. The linguistic capacities children acquire, and the routes by which these capacities are reached, admit of no other explanation.

This is not to say that linguistic nativists are all agreed on the details of our innate language-specific components. Mark Baker (2005), for example, points out that while it is clear that something like the Head Directionality Principle exists as part of our innate linguistic endowment, it is not yet clear whether this principle is innately specified as:

1. Combine a word with a phrase to make a larger phrase by putting the new word first *or* by putting the new word last.

Or as:

2. Form a new phrase by adding a word to a phrase.

More generally, there is still much disagreement between linguistic nativists over whether our innate linguistic endowment is overspecified (as in 1) or underspecified (as in 2). Chomsky and many linguists who follow him currently argue broadly in favor of

the former, whereas Steven Pinker, most ‘functional’ linguists, and most developmental psychologists argued broadly for the latter. Resolution of this debate will ultimately be an empirical matter, of course, but current theoretical arguments do not tell decisively either way. Nor is it clear how similar issues in other cognitive domains should either influence or be influenced by results from language acquisition. Thus despite the overwhelming nature of the theoretical arguments and empirical data outlined in the previous paragraphs in favor of linguistic (and other) nativism, there still exist some significant – though healthy – disagreements between even the most canonical nativists over exactly what the details of our innate endowment are. Nonetheless, such disagreements should not obscure the fact that, in the linguistic domain and elsewhere, our innate endowment is undeniably both rich and complex.

Finally, there is also the question of how our (neuro-)biology is able to implement or otherwise provide the innate knowledge which our linguistic and other cognitive abilities require. Given that aspects of our linguistic competence are not the product of our linguistic experience, it seems that these aspects must therefore be part of our biological endowment. This may strike one as somewhat counter-intuitive, but as Stephen Crain, Andrea Gualmini, and Paul Pietroski (2005) have recently argued “[o]ne cannot insist that our shared biology cannot give rise to knowledge of specific contingent linguistic facts if the available evidence suggests that our shared biology does just this.” Moreover, it may turn out that the nature of these innate linguistic facts actually provides us with useful information about the manner in which human biology has evolved, given the evolutionary pressures and constraints which (proto-)language use will have imposed upon the primate lineage. Indeed, investigation into the interplay between our innate linguistic knowledge and the evolution of our species-typical biology, will undoubtedly be one of the more complex and fruitful areas for future research. The same is also likely to be true with regard our investigation into and our understanding of the innate knowledge involved in other cognitive domains.

In sum, then, in one important sense of ‘innate knowledge’ the existence of such knowledge in the context of linguistic and various other cognitive capacities is no longer in any serious doubt. However, much work is still required before our understanding of the specific nature of this knowledge will be complete.

See also: Concepts; Innate Ideas; Modularity; Plato and His Predecessors.

Bibliography

- Ariew A (1999). 'Innateness is canalisation.' In Hardcastle V (ed.) *Where biology meets psychology: philosophical essays*. Cambridge, MA: MIT Press. 117–138.
- Baker M (2005). 'The innate endowment for language: overspecified or underspecified?' In Carruthers P, Laurence S & Stich S (eds.) *The innate mind: structure and content*. Oxford: Oxford University Press.
- Barkow J, Cosmides L & Tooby J (eds.) (1992). *The adapted mind*. New York: Oxford University Press.
- Botterill G & Carruthers P (1999). *The philosophy of psychology*. Cambridge: Cambridge University Press.
- Brown D (1991). *Human universals*. New York: McGraw-Hill.
- Carruthers P, Laurence S & Stich S (2005). *The innate mind: structure and content*. Oxford: Oxford University Press.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky N (1967). 'Recent contributions to the theory of innate ideas.' *Synthese* 17, 2–11.
- Chomsky N (1980). *Rules and representation*. New York: Columbia University Press.
- Chomsky N (1981). *Lectures on government and binding*. Holland: Foris.
- Crain S, Gualmini A & Pietroski P (2005). 'Brass tacks in linguistic theory: innate grammatical principles.' In Carruthers P, Laurence S & Stich S (eds.) *The innate mind: structure and content*. Oxford: Oxford University Press.
- Crain S & Pietroski P (2001). 'Nature, nurture and Universal Grammar.' *Linguistics and Philosophy* 24, 139–186.
- Fodor J (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Fodor J (2000). *The mind doesn't work that way*. Cambridge, MA: MIT Press.
- Griffiths P (2002). 'What is innateness?' *The Monist* 85, 70–85.
- Heyes C & Huber L (eds.) (2000). *The evolution of cognition*. Cambridge, MA: MIT Press.
- Karmiloff-Smith A (1992). *Beyond modularity*. Cambridge, MA: MIT Press.
- Laurence S & Margolis E (2001). 'The poverty of the stimulus argument.' *British Journal of Philosophy* 52, 217–276.
- Locke J (1690). *An essay concerning human understanding*. In Nidditch P (ed.) (1975). Oxford: Clarendon Press.
- Lorenz K (1957). 'The nature of instincts.' In Schiller C (ed.) *Instinctive behavior*. New York: International University Press. 129–175.
- Marcus G (1993). 'Negative evidence in language acquisition.' *Cognition* 46, 53–85.
- Pinker S (1994). *The language instinct*. New York: William Morrow.
- Pinker S (1997). *How the mind works*. London: Allen Lane.
- Pinker S (2002). *The blank slate*. London: Allen Lane.
- Plato (1997). *The dialogues of Plato*. Jowett B (ed.) Routledge: Thoemmes Press.
- Pye C (1992). 'The acquisition of K'iché Maya.' In Slobin D (ed.) *The crosslinguistic study of language acquisition*, vol. 3. Hillsdale, NJ: Earlbaum. 221–308.
- Samuels R (2000). 'Massively modular minds: evolutionary psychology and cognitive architecture.' In Carruthers P & Chamberlain A (eds.) *Evolution and the human mind: modularity, language and meta-cognition*. Cambridge: Cambridge University Press. 13–46.
- Samuels R (2002). 'Nativism in cognitive science.' *Mind and Language* 17, 233–265.
- Stich S (1975). 'Introduction: The idea of innateness.' In Stich S (ed.) *Innate ideas*. Berkeley: University of California Press. 1–22.
- Stromsworld K (2000). 'The cognitive neuroscience of language acquisition.' In Gazzaniga M (ed.) *The cognitive neurosciences*. Cambridge, MA: MIT Press. 909–932.

Intention and Semantics

S Barker, University of Nottingham, Nottingham, UK

© 2006 Elsevier Ltd. All rights reserved.

A compelling idea is that words have meanings because speakers who use them have intentions of some kind. Intentions underpin semantics. One proposal of how this might work is based on Locke's idea (1975) that the function of a language is to express preexisting thoughts; thoughts are correlated with sentences through intentions. So intentions are the glue binding thoughts to words.

Grice (1958, 1969) adopts this Lockean view. Grice divides verbal meaning into two kinds: so called 'speaker-meaning' and 'sentence-meaning.' Speaker-meanings are the contents of particular utterances on occasions. They are the contents of illocutionary acts (in the sense of Austin, 1962) performed by production of sentences. Illocutionary acts include assertions, orders, questions, etc. Speaker-meanings may diverge from sentence-meanings, which are literal meanings. U may employ 'The church is a rock' but convey as speaker-meaning that the church is a strong support for its members; U may utter 'The meal is

edible,' conveying as speaker-meaning that it is not very good.

Grice begins his analysis of meaning by examining the intentions underpinning speaker-meaning. Grice takes the intentions that do the job to be communicative intentions. A communicative intention is social: it is directed toward an audience. In communication – speaking-meaning something – a subject *U* intends that her audience *H* gain a certain state *r*, a belief, desire, intention, etc. Grice's great insight was into the structure of the communicative intention, which he characterized thus:

C: *U* means something by utterance *S* if and only if *U* utters *S* and intends that *H* acquire characteristic *r* and *U* intends that *H* do so partly by recognizing this very intention.

The intention is a **reflexive intention**: an intention that falls within its own scope. Many philosophers have been wary of the reflexive treatment of communication and have attempted to explicate communicative intentions in terms of hierarchies of intentions. See Grice (1969), Strawson (1964), and Schiffer (1972). But the results are unwieldy.

Speaker-meanings are, then, acts with the form C. The different types of speaker-meanings or illocutionary acts, that is, assertions and orders etc., are determined by the different characteristics *r* intended. Grice (1971: 123) analyzes assertions and orders thus:

GA: *U* asserts (to *H*) that *P* by uttering *S* iff *U* utters *S* reflexively intending that *H* believe that *U* believes that *P*.

GO: *U* orders (*H*) to do *F* by uttering *S* iff *U* utters *S* reflexively intending that *H* form an intention to *F*.

Such proposals about *r* are problematic, however. Restricting ourselves to assertion, GA is refuted by apparent cases in which *U* lacks an intention that *H* believe that *P* – e.g., where *U* either (*a*) is indifferent as to *H*'s belief, because, for example, *U* is engaged in polite conversation without intending to persuade; or (*b*) believes *H* won't believe her. (See Alston, 2000). Bach and Harnish (1979), suggest the primary intention should rather be:

that *H* has a reason to believe that *U* believes that *P*.

If, as suggests Recanati (1986), reasons are defeasible, then *U* can provide a reason for *H* to believe *P*, even though *U* knows it is undermined by further information, such as that *H* believes *U* is a liar.

Assuming that some such explication of *r* works, how can we use speaker-meaning to analyze sentence-meaning? Grice (1971) introduces practices and conventions thus:

SM1: *S* means that *P* for *U* iff *U* has the practice that if she desires to reflexively intend that *H* gain *r*, then she (may) utter *S*.

The 'may' here can be taken as rule permissibility or epistemic 'may.' SM1 allows words to be ambiguous, and for there to be different ways of saying the same thing. Grice notes that the regularity governing use of the sentence *S* does not itself have to be one that correlates use of *S* with full-fledged reflexive intentions. Rather we simply need:

SM2: *S* means that *P* for *U* iff *U* has the practice that if she desires to intend that *H* gain *r*, then she (may) utter *S*.

If the conventions/practices governing sentences have this form, we can then explain how speaker-meanings – full illocutionary acts – emerge in particular utterance events. In uttering *S*, *U* intends *H* to think thus:

U's policy is that if *U* intends that *H* have *r*, then *U* (may) utter *S*.

U uttered *S*, so (given context) I can infer that *U* intends that *H* have *r*.

U intends that *H* have *r*.

So, I should have *r*.

Thus, *U* is intending that *H* have *r*, partly in virtue of recognizing *U*'s intention that she have *r*. But this is just reflexively intending that *H* have *r*.

To analyze word-meaning we need regularities such as:

WM: The word Ω means *O* for *U* iff *U* has the practice that if *U* desires that *H* believe *U* believes something about *O*, then *U* (may) utter a sentence ... Ω ...

Given that *U* has such dispositions for the basic vocabulary of her language, we can deduce the kinds of dispositions that will underpin her production of novel (unuttered) sentences. We provide thereby a compositional account of the meaning of sentences in *U*'s language.

Platts (1979: 86–94) doubts this last point. He thinks the very thing we are meant to be explicating, sentence-meaning, will have to be brought in to fix the intentions that *U* would have if she were to use novel sentences. Blackburn (1984: 127–129) objects that it is just the practices and dispositions themselves that fix what we would intend, and thus mean. Nothing else need back up such dispositions.

A more serious objection to the Gricean account of meaning is that repertoire rules for the words require us to appeal to the semantic relation of **aboutness**. WM invokes *U*'s believing something **about** object *O*. Aboutness is related to denotation, that is, a representational relation between mental state or word

and world. But meanings, one might think, are simply representations. So, as Devitt and Sterelny (1999: 150–151) point out, the Gricean analysis effectively leaves unexplained the realm of content as such. Perhaps we need a Fodorean (1975) language of thought hypothesis to finish the story. Or perhaps there is some way of giving a pragmatic reduction of representation in terms of an inferential semantics (Brandom, 1994). Or we should see Grice's analysis of speaker-meaning and its relation to sentence-meaning as merely providing an explanation of what it is for particular speakers to use one language – whose representational contents are abstractly defined in terms of truth-conditions – rather than another (Lewis, 1975). If so, as Blackburn (1984: 134) points out, there is no rivalry between intention-based semantics and formal truth-conditional approaches, *contra* Strawson (1964).

Or there is this possibility. An intention-based semantics might change tack by denying that meanings, *qua* semantic interpretations, are representations as such. This approach is found in Barker (2004). It contends that the semantic interpretations of words and sentences are not representational contents but speech act types – acts whose nature is not merely representational. These speech acts are a specific kind called 'proto-speech-acts.' A proto-act is uttering a word, phrase, or sentence and **advertising** certain intentions to denote, represent, or communicate. Advertising is engaging in the behavior characteristic of a speaker who, following certain rules, intends to do something. In uttering a name, for example *Pegasus*, U utters a term, *Pegasus*, and advertises an intention to denote an object, something called *Pegasus*. If U utters 'I saw Pegasus,' she wants H to believe she has the intention she advertises, but if she asserts 'Pegasus does not exist,' she does not. What defines a name's meaning is not any object denoted, since like *Pegasus* it might be empty, but the proto-referring act associated with it. The meaning of *Pegasus* is that proto-referring act all of whose tokens are nodes of a certain referential tree: that subset of uses of the name *Pegasus* that we group together as instances of the name for the mythical flying horse.

The meaning of a declarative sentence is a 'proto-assertion.' A proto-assertion involves two parts: advertising (a) a representational intention and (b) a communicative intention. In uttering 'Snow is white,' U advertises intentions (a) to represent that snow is white, and (b) to **defend** the intention to represent snow's whiteness. Defending is dialectical engagement with an audience H: in defending an intention to represent that snow is white, one wants H to accept or reject such an intention in her own

case, and to give reasons for rejection. H judges correct an **assertion** of 'Snow is white' – where U really does intend to defend an intention to represent that snow is white – iff H accepts what U defends.

Assertion of 'Snow is white' is a **report**, since what is defended is a representational intention. But not all assertions are reports. Utterance of 'Haggis is tasty' is an expression of taste: U advertises intentions to (a) represent her possession of a gustatory preference state of liking haggis, and to (b) defend that gustatory preference state. Preference states are not representational. H judges correct U's assertion iff H accepts the gustatory preference state in her own case; not if and only if H thinks U has represented correctly her, U's, state.

In this framework, all logically complex sentences have expressive proto-assertions as their meanings. Negations express rejective states. In uttering 'Snow is not white,' U advertises intentions to (a) represent that she has a rejective state with respect to intending to represent snow is white, and to (b) defend that rejective state. In uttering 'Haggis is not tasty,' U expresses rejection of the gustatory property. And so on for other logically complex sentences. Statements of the form 'S is true' are expressive as well; U expresses her acceptance of the state defended in assertion of S. Because sentence-meanings are not representations, we are not committed to logically complex entities in the world, such as negative or universal facts, or mysterious properties of truth.

A compositional semantics can be built through constructing meanings in terms of proto-acts, proto-referrings, and proto-assertions. Proto-assertions, for example, can embed in logical compounds. In 'either S or R,' S and R are proto-asserted. Hence **advertising** intentions is weaker than the condition articulated above of giving a defeasible reason to believe.

This approach does not attempt to explicate representation in speech act terms. Rather, it displaces representation as the keystone of meaning. Names and sentences don't have to denote/represent to be meaningful. Truth-bearers are not propositions, *qua* representational contents, but assertions: acts of defending states. This account does not tell us what denotation and representation are, but, unlike the Gricean approach, it is not committed to saying that meaning resides in such relations holding. The result is an intention-based semantics that seriously challenges the dominant truth-conditional approach to meaning.

See also: Assertion; Compositionality: Philosophical Aspects; Expression Meaning versus Utterance/Speaker Meaning; Ideational Theories of Meaning; Thought and Language: Philosophical Aspects; Truth Conditional Semantics and Meaning.

Bibliography

- Alston W J (2000). *Illocutionary acts and sentence meaning*. Syracuse: Cornell.
- Austin J (1962). *How to do things with words*. Urmson J O & Warnock G J (eds.). Oxford: Oxford University Press.
- Bach K & Harnish R M (1979). *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.
- Barker S J (2004). *Renewing meaning*. Oxford: Clarendon Press.
- Blackburn S J (1984). *Spreading the word*. Oxford: Clarendon Press.
- Brandom R (1994). *Making it explicit*. Cambridge, MA: Harvard University Press.
- Devitt M & Sterelny K (1999). *Meaning and reality*. Cambridge, MA: MIT Press.
- Fodor J (1975). *The language of thought*. New York: Crowell.
- Grice P (1958). 'Meaning.' *Philosophical Review* 67, 377–388.
- Grice P (1969). 'Utterer's meaning and intentions.' *Philosophical Review* 78, 147–177.
- Grice P (1971). 'Utterer's meaning, sentence-meaning, and word-meaning.' In Searle J (ed.) *The philosophy of language*. Oxford: Oxford University Press.
- Lewis D K (1975). 'Languages and language.' In Gunderson K (ed.) *Minnesota studies in the philosophy of science VII, Book 3, ch. 1–3*. Minneapolis: University of Minnesota Press.
- Locke J (1975). *An essay concerning human understanding*. Nidditch P (ed.). Oxford: Oxford University Press.
- Platts M (1979). *The ways of meaning: an introduction to a philosophy of language*. London: Routledge & Kegan Paul.
- Recanati F (1986). 'On defining communicative intentions.' *Mind and Language* 1, 213–241.
- Schiffer S (1972). *Meaning*. Oxford: Clarendon Press.
- Strawson P (1964). 'Intention and convention in speech acts.' *Philosophical Review* 73, 439–460.

Interpreted Logical Forms

M Montague, University of California, Irvine, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

Interpreted logical forms (ILFs) were originally introduced by Harman (1972) to answer the question: What are the objects of propositional attitudes (belief, desire, hope, regret, etc.)? The theory has since been developed by a number of philosophers, most notably Higginbotham (1986, 1991), Segal (1989), Larson and Ludlow (1997), and Larson and Segal (1995). Seymour (1996) has suggested that ILF theories can also solve certain puzzles in quotational environments.

Propositional Attitude Reports

Consider the following two propositional attitudes:

- (1) Lois believes that Superman can fly.
- (2) Lois believes that Clark Kent can fly.

It is an 'intentional fact' that Lois may have the attitude displayed in (1) without having the attitude displayed in (2). It is a 'semantic fact' that sentences reporting these propositional attitudes may differ in truth value. Are these facts the same? No: one is a fact about intentionality, and the other is a fact about semantics. Is this difference important? According to the theory of ILFs advanced by Ludlow, Larson, and Segal, it is. Giving a semantics for

propositional-attitude reports, which is the goal of the ILF theories, has virtually nothing to say about propositional attitudes themselves: "The ILF theory ... (as a semantic theory) ... addresses only the truth conditions of sentences involving *believe*, *think*, *assert*, etc., it does not address the beliefs, thoughts, and assertions of persons" (Larson and Ludlow, 1997: 1035). By contrast, the ILF theory proposed by Higginbotham (1986, 1991) countenances a closer relationship between the theory of propositional attitudes and the semantics of attitude reports.

Since Frege (1892), a huge amount of attention has been given to the aforementioned semantic fact, which has become known as 'Frege's Puzzle.' Frege discovered that if 'Superman' and 'Clark Kent' refer to their ordinary referents in (1) and (2), according to the principle of substitution, that co-referring expressions may be substituted *salva veritate*, (1) and (2) should share a truth value. Since they intuitively do not, the principle of substitution seems to fail in propositional-attitude contexts. Frege's puzzle has given rise to the very difficult project of giving a semantic theory that deals satisfactorily with propositional-attitude reports.

Frege offered a solution based on his sense/reference distinction. In addition to extensional entities (individuals, sets, and relations-in-extension) he postulated senses or modes of presentation, which we now sometimes call intensions, and which are, roughly, ways of determining referents. Briefly, in

attitude contexts, ‘Superman’ and ‘Clark Kent’ do not refer to their ordinary referents, but to their ordinary senses. Since ‘Superman’ and ‘Clark Kent’ refer to different senses in (1) and (2), we shouldn’t expect to substitute them *salva veritate*. In this way Frege preserves the principle of substitution, but at a cost. First, he introduces intensions, which seem to many to be dubious entities (Quine, 1961). Second, he violates Davidson’s (1968) requirement of ‘semantic innocence’ – that expressions should have the same semantic values in all contexts.

ILF theories aim to offer a semantic theory of propositional-attitude reports that can avoid these supposed problems. They aim to give a purely extensional semantics for attitude contexts, while preserving semantic innocence. Achieving these aims would make them extremely attractive. (See Larson and Segal, 1995: 437 for their rejection of intensional strategies for preserving semantic innocence.)

ILF theories are standardly embedded in a Davidsonian truth-theoretic semantics for natural language. The introduction of ILFs allows one to provide truth conditions for sentential complements embedded in propositional-attitude reports **within** the truth-theoretic-semantics framework. (For a clear explanation of why the Davidsonian framework (without ILFs) is not adequate to capture the truth conditions of propositional-attitude reports, see Larson and Segal, 1995: 415–418.)

What Are ILFs?

The basic idea is that propositional-attitude verbs express relations between an agent and an ‘interpreted logical form.’ The ‘logical form’ part of an ILF is, roughly, a sentential complement (a syntactic item), which usually takes one of the following forms: ‘a is F,’ ‘a’s being F,’ or ‘that a is F.’ (Lexical items are part of an expression’s logical form.) The ‘interpretation’ part of an ILF is the assignment of referents to parts of the sentential complement. ILFs can be represented with phrase structure trees whose terminal nodes are pairings of the relevant lexical items with their referents, and whose nonterminal nodes are pairings of the relevant phrasal categories with their referents. Simplifying, the sentential complement in an attitude reports such as

- (3) Lois believes John is funny

receives the following ILF:

<S, True>	
<NP, John>	<VP, John>
<‘John’, John>	<‘Funny’, John>

In a nutshell, ILFs are a combination of syntactical (linguistic) and referential (nonlinguistic) material – they are hybrids relative to purely syntactical approaches and purely referential approaches. (See Larson and Segal, 1995: 419–422 for a discussion of purely syntactical approaches.) Since ILFs conjoin only lexical items and extensional referents, ILF theories provide a purely extensional semantics for propositional-attitude reports.

Puzzles and Problems

An adequacy test for any semantics of the attitudes is how well it does in solving traditional puzzles. To begin, consider the way in which the linguistic and nonlinguistic features of ILFs allow them to solve one puzzle involving names and one puzzle involving demonstratives. I call these the ‘simple name puzzle’ and the ‘simple demonstrative puzzle.’ (The use of the term ‘simple’ here is of course theory relative. Russellians (e.g. Salmon, 1986, and Soames, 1987) have great difficulty with name puzzles.)

The Simple Name Puzzle

Most agree that the following two belief reports can differ in truth value, despite the fact that ‘Fido’ and ‘Rex’ refer to the same dog (this is a variant of the aforementioned Clark Kent/Superman case):

- (4) John believes that Fido barks.
(5) John believes that Rex barks.

John, for example, may only know the dog by the name ‘Fido’ and so only believe (4). Because ILFs are partly constituted by lexical items, the ILF theory can easily account for the difference in truth value between (4) and (5). Since ‘Fido’ and ‘Rex’ are different lexical items, the ILFs associated with the sentential complements in (4) and (5) will be different.

The Simple Demonstrative Puzzle

Consider

- (6) John believes that that is funny

used to report John’s referring to a comedian’s skit with the use of the embedded demonstrative. Now consider (6) used to report John’s referring to a TV show with the embedded demonstrative: clearly a different belief report. A purely sententialist theory of the attitudes (one that appeals only to syntactical items) has difficulty accounting for this difference. In both instances, John would be standing in relation to the same sentential complement, ‘that is funny.’ Since ILFs contain the referents of lexical items, they can easily account for differences between belief reports

that share sentential complements. In this case, one ILF would contain the comedian's skit as the semantic value of the embedded 'that,' and the other would contain the TV show.

These two puzzles show the power of ILF theories. The lexical features of ILFs allow them to be fine-grained enough to solve simple name puzzles without postulating intensional entities. The objectual features of ILFs prevent them from being too coarse-grained (as purely syntactical theories are) to solve simple demonstrative puzzles.

Consider now two further puzzles that prove to be more difficult for ILF theories. I call these 'the hard name puzzle' and 'the hard demonstrative puzzle.'

The Hard Demonstrative Puzzle

Suppose that John assents to

(7) That is a philosopher

while pointing to his professor in philosophy class (context 1), but denies (7) while pointing to a man seen from behind (he doesn't know it's his professor) at a party (context 2). Relative to these different contexts, then, intuitively,

(8) John believes that that is a philosopher

according to the context in which it is uttered, can differ in truth value. On its face, since the sentential complement, 'that is a philosopher,' and the referents assigned to its parts are the same in both contexts, ILF theories appear unable to account for the possible difference in truth value.

Larson and Ludlow (1997) respond to the hard demonstrative puzzle by appealing to a Burgean (1974) idea about the semantic values of demonstratives: according to Burge, the semantic value of a demonstrative is a pairing of what is demonstrated and the act of demonstration. By including acts of demonstration in the semantic value of demonstratives in attitude contexts, Larson and Ludlow can account for the truth value differences of (8) since the acts of demonstration will be different in the two contexts. Relative to context 1 the semantic value of the demonstrative will be $\langle x, a1 \rangle$, where x is the philosopher and $a1$ is the speaker's demonstrative act, and relative to context 2 the semantic value will be $\langle x, a2 \rangle$ where $a2$ is a distinct demonstrative act from $a1$.

Pietroski (1996) argues that it is unclear whether this response is nonFregean and well motivated. That is, Pietroski observes that acts of demonstration can affect the truth of (8) but not the truth of (7). This, he claims, concedes the Fregean point that a demonstrative way of thinking of a referent affects the truth

value of a sentence with an **embedded** demonstrative. Pietroski suggests that Larson and Ludlow can avoid this complaint by agreeing with Burge (1974), on grounds independent of attitude contexts, that the semantic values of demonstratives are **always** ordered pairs of the thing demonstrated and the act of demonstration.

A suggestion made by Higginbotham (1991) may provide an alternative way of dealing with cases like (8). In discussing an objection to both Creswell's (1985) structured meanings account of the attitudes and Larson and Ludlow's theory of ILFs, Higginbotham argues that both fail to include a crucial parameter in their accounts: "... namely that complement sentences are to be understood as if their speakers said them" (1991: 352). Lau (1995) argues that applying this parameter to (7), relative to context 1, in uttering (7) John would understand himself as referring to his philosophy professor and (8) would turn out true, but relative to context 2, John would not understand himself as referring to his philosophy professor and (8) would turn out false. Although this suggestion accounts for this case, more would need to be said about how this new parameter should be incorporated into the semantics.

The Hard Name Puzzle

Kripke's (1997) Paderewski case (and others like it) pose a more difficult challenge for ILF theories. Suppose John comes to know his shy neighbor, who never plays music at home and does not own a piano, as 'Paderewski.' Thus (9) is true:

(9) John does not believe that Paderewski is a pianist.

John also attends many concerts where he meets his favorite pianist, also called 'Paderewski.' Although these are the same Paderewski, John does not realize this. So, it is also true that

(10) John believes that Paderewski is a pianist.

There seems to be no way of distinguishing John's beliefs either in terms of lexical items (since there is only one linguistic item, 'Paderewski') or in terms of semantic values (since there is only one referent, Paderewski). So ILF theories apparently attribute contradictory beliefs to John.

The different responses to this puzzle (and the complexity of some of them) are evidence that it is particularly difficult for ILF theories. Larson and Ludlow (1993) suggest that although (9) and (10) contain the homophonous 'Paderewski,' there are really distinct syntactical items in play – 'PaderewskiI' and 'PaderewskiII' (much as there are distinct

syntactical items associated with 'bank'). They argue that since there are actually two names in (9) and (10), different ILFs would be assigned to them, thus avoiding the assignment of contradictory beliefs to John.

Many (e.g., Forbes, 1996; Taylor, 1995; Richard, 1990) have found this response implausible. They argue that unlike the term 'bank,' there is only one name 'Paderewski' in our common language. John certainly doesn't seem to have had a private baptism, introducing two names into his idiolect; rather he picked up the name in the ordinary way. Since he picked up the name in the ordinary way, and our language only has one name, John is using only one name. (See Pietroski, 1996: 366–368 for a different objection to Larson and Ludlow's solution to the Paderewski puzzle.)

Ludlow (2000) attempts a different response to the Paderewski puzzle, arguing that it dissolves on a correct understanding of language and the goals of ascribing propositional attitudes. First, language is not an external, social object, but what Chomsky (1986) calls an I-language – each person has her own lexicon, relying on substantial overlap with others for communication. Second, in ascribing propositional attitudes we are not trying to describe what is going on in an agent's head. Rather, we are helping a hearer construct a theory of an agent's mental life. Our focus, therefore, should be on the relationship between the ascription and the hearer, not on the ascription and the bearer of the attitude. Working with the hearer is a complicated process that involves theories of tacit belief, of goals of belief ascription, and of belief ascription logistics. But most importantly for Paderewski cases, the speaker and hearer are engaged in a negotiation about the best expressions for ascribing a propositional attitude to the agent. In Paderewski cases we need some way of distinguishing between an agent's beliefs, and the speaker and hearer negotiate this by choosing distinct expressions. Signaling John's beliefs about Paderewski *qua* piano player will involve an appropriate choice of expression – perhaps speaker and hearer will use the expression 'Paderewski *qua* piano player.' Similarly for John's beliefs about Paderewski *qua* shy neighbor. The upshot is that we do not get a straightforward contradiction of the form (Fa & ~Fa) – our different I-languages grant us fluidity with our lexicons, making contradictions across discourses unlikely.

Prospects

Interestingly, this response to the Paderewski case is related to another common complaint about ILF theories. Maintaining that ILFs are partly constituted

by English lexical items results in several related difficulties: if a monolingual French speaker believes that Fido barks, he does not believe an ILF with English expressions; ILF theories seem incapable of capturing the idea that English and French speakers believe the same thing (or something very similar) when they believe that Fido barks; phonetically distinct sentential complements cannot co-refer. These problems result from the plausible idea that a semantic theory ought to capture what is grasped in understanding sentences of a language; and speakers using different languages can presumably grasp the same thing. This immediately forges a close link between the objects of the attitudes and the semantics of attitude ascriptions.

These problems may be dealt with semantically or pragmatically. Davidson (1968), Lepore and Loewer (1989) and Higginbotham (1986) suggest a semantic solution by building a notion of same-saying or similarity into the truth conditions for attitude reports. Agents stand in intentional relations to ILFs similar to (or that say the same thing as) those involving English lexical items. The similarity or same-saying relation provides a way of capturing what is common among thinkers across languages and thinkers without linguistic abilities. This is accomplished by indicating a close relation between the theory of propositional attitudes and the semantics of attitude ascriptions.

Larson and Ludlow (1997), Larson and Segal (1995), and Larson (2000) offer a pragmatic solution. Similarity of propositional attitude or same-saying is a matter of usage, not content: it is a pragmatic matter whether two propositional-attitude sentences can be used to report the same attitude. The form of this pragmatic solution has been cast in different ways. In Larson and Ludlow (1997), choosing the best propositional-attitude sentence is based on the same considerations Ludlow (2000) appealed to in addressing the Paderewski puzzle: speaker and hearer work out the best attitude ascriptions based on theories of tacit belief, goals of belief ascription, and belief ascription logistics. Larson and Segal (1995) rely on the notion of **expression** to elucidate their pragmatic solution. In short, when an ILF is used to report an agent's belief, it is not that an agent is standing in a relation to an ILF; rather, the ILF is **used to express** what the agent believes: "to believe an ILF is to have a belief expressed by it, to desire an ILF is to have a desire expressed by it, and so on" (Larson and Segal, 1995: 445). Explicating believing and desiring in terms of the notion of expression is not part of the semantics – it is an explanation of what believing and desiring an ILF involves. So, to believe that grass is green is to **believe an ILF**, which means to have a belief that is expressed by

that ILF. But what does an ILF express? The answer to this question seems to present ILF theorists with a dilemma. Either an ILF expresses a proposition, something that transcends any specific language, or an ILF expresses something that involves English expressions. The first horn of the dilemma seems to lead in the direction of Fregean propositions, and the second horn of the dilemma does not seem capable of capturing what a French speaker believes.

In summary, ILFs are combinations of linguistic items and nonlinguistic items. Their linguistic features are at once their most coveted and their most objectionable feature. They provide a solution to a traditional name puzzle without postulating intensional entities. But if the semantics of attitude reports indicate what we stand in relation to when we believe, desire, and hope, it is problematic to appeal to English expressions. If one chooses a semantic solution to this problem, building a similarity or same-saying relation into the truth conditions, then a careful consideration of these relations is required. If one chooses a pragmatic solution, one must decide whether one has given up too much in surrendering the idea that the semantics of attitude reports overlaps in a natural way with the theory of the attitudes.

See also: Propositional Attitude Ascription: Philosophical Aspects; Representation in Language and Mind; Semantics–Pragmatics Boundary; Sense and Reference: Philosophical Aspects; Thought and Language: Philosophical Aspects.

Bibliography

- Bealer G (1993). 'A solution to Frege's puzzle.' *Philosophical Perspectives, Language and Logic* 7, 17–60.
- Burge T (1974). 'Demonstrative constructions, reference and truth.' *The Journal of Philosophy* 71, 205–223.
- Chomsky N (1986). *Knowledge of language*. New York: Praeger.
- Crane T (2001). *Elements of mind*. Oxford: Oxford University Press.
- Creswell M J (1985). *Structured meanings*. Cambridge, MA: MIT Press.
- Davidson D (1968). 'On saying that.' In Davidson D (ed.) *Inquiries into truth & interpretation*. Oxford: Clarendon Press. 93–108.
- Den Dikken M, Larson R & Ludlow P (1997). 'Intensional "transitive" verbs and concealed complement clauses.' In Ludlow P (ed.).
- Fiengo R & May R (1996). 'Interpreted logical form: a critique.' *Rivista di Linguistica* 105, 349–374.
- Forbes G (1990). 'The indispensability of Sinn.' *Philosophical Review* 99, 535–563.
- Forbes G (1996). 'Substitutivity and the coherence of quantifying in.' *Philosophical Review* 105, 337–372.
- Forbes G (2000a). 'Intensional transitive verbs: the limitations of a clausal analysis.' Unpublished ms., <http://www.Tulane.edu/~forbes/preprints.html>.
- Forbes G (2000b). 'Objectual Attitudes.' *Linguistics and Philosophy* 23, 141–183.
- Forbes G (2002). 'Intensionality.' *Proceedings of the Aristotelian Society supplementary volume* 76, 75–99.
- Frege G (1892). 'On sense and nominatum.' In Martinich A P (ed.) *The philosophy of language*, 3rd edn. Oxford: Oxford University Press. 186–198.
- Harman G (1972). 'Logical Form.' *Foundations of Language* 9, 38–65.
- Higginbotham J (1986). 'Linguistic theory and Davidson's program in semantics.' In Lepore E (ed.) *Truth and interpretation: perspectives on the philosophy of Donald Davidson*. Oxford: Basil Blackwell. 29–48.
- Higginbotham (1991). 'Belief and logical form.' *Mind & Language*, 344–369.
- Kripke S (1997). 'A puzzle about belief.' In Ludlow P (ed.). 875–920.
- Larson R (2002). 'The grammar of intensionality.' In Peter G & Preyer G (eds.) *Logical form and language*. Oxford: Clarendon Press. 228–262.
- Larson R & Ludlow P (1997). 'Interpreted logical forms.' In Ludlow P (ed.). 993–1039.
- Larson R & Segal G (1995). *Knowledge of meaning*. Cambridge: MIT Press.
- Lau J (1995). 'Belief reports and interpreted-logical forms.' Unpublished manuscript, <http://www.hku.hk/philodep/joelau/phil/ILF.htm>.
- Lepore E & Loewer B (1989). 'You can say *that* again.' In French P, Uehling T & Wettstein H (eds.) *Midwest Studies in Philosophy XIV: Contemporary Perspectives in the Philosophy of Language II*. Notre Dame: University of Notre Dame Press.
- Ludlow P (ed.) (1997). *Readings in the philosophy of language*. Cambridge: MIT Press.
- Ludlow P (2000). 'Interpreted logical forms, belief attribution, and the dynamic lexicon.' In Jaszczolt K M (ed.) *The pragmatics of propositional attitude reports*. Oxford: Elsevier Science.
- Pietroski P (1996). 'Fregean innocence.' *Mind & Language* 11, 338–370.
- Quine W O (1961). *From a logical point of view*. New York: Harper and Row.
- Richard M (1990). *Propositional attitudes*. Cambridge: Cambridge University Press.
- Salmon N (1986). *Frege's puzzle*. Cambridge: MIT Press.
- Segal G (1989). 'A preference for sense and reference.' *The Journal of Philosophy* 86, 73–89.
- Seymour D (1996). 'Content and quotation.' *Rivista di Linguistica* 8, 309–330.
- Soames S (1987). 'Direct reference, propositional attitudes, and semantic content.' *Philosophical Topics* 15, 47–87.
- Taylor K (1995). 'Meaning, reference, and cognitive significance.' *Mind & Language* 10, 129–180.

Irony

S Attardo, Youngstown State University, Youngstown, OH, USA

© 2006 Elsevier Ltd. All rights reserved.

The term 'irony' is commonly used to describe both a linguistic phenomenon (verbal irony) and other phenomena including 'situational' irony (i.e., irony of facts and things dissociated from their linguistic expression; Shelley, 2001) such as a fire-station burning to the ground, various more-or-less philosophical ideas (Socratic irony, Romantic irony, Postmodern irony), and even a type of religious experience (Kierkegaard, 1966). While there may be connections between situational and verbal irony, it does not appear that literary and religious uses can be fruitfully explained in terms of linguistic irony. This treatment will be limited to verbal irony.

Other definitional problems include the purported distinction between irony and sarcasm. While some have argued that the two can be distinguished (for example, irony can be involuntary, while sarcasm cannot be so), others maintain that no clear boundary exists. A further problem is presented by the fact that in some varieties of English, the term *irony* is undergoing semantic change and is assuming the meaning of an unpleasant surprise, while the semantic space previously occupied by *irony* is taken up by the term *sarcasm*.

The word *irony* goes back to the Greek *eironeia* (pretense, dissimulation) as does the history of its definition and analysis. Irony is seen as a trope (i.e., a figure of speech) in ancient rhetorics and this analysis has remained essentially unchallenged until recently. In the traditional definition irony is seen as saying something to mean the opposite of what is said. This definition is demonstrably incorrect, as a speaker may be ironical but not mean the opposite of what he/she says; cf. *It seems to be a little windy* (uttered in the middle of a violent storm), in which the speaker is saying less than what is meant. Similarly, overstatements and hyperbole may be ironical (Kreuz and Roberts, 1995).

A recent and fruitful restatement of the irony-as-trope theory has been presented by Paul Grice who sees irony as an implicature, i.e., as a deliberate flouting of one of the maxims of the principle of cooperation. Relatedly, speech-act approaches to irony see it as an insincere speech act. Initially, Grice's approach saw irony as a violation of the maxim of quality (i.e., the statement of an untruth) but this claim has been refuted, as seen above. Broadening the definition to, for example, 'saying something while meaning

something else,' runs the risk of obliterating the difference between irony and other forms of figurative or indirect speech. However, this loss of distinction may be a positive aspect of the definition, as has been recently argued (Kreuz, 2000, Attardo, 2002).

While the idea of 'oppositeness' in irony is problematic, approaches to irony as negation have been presented (Giora, 1995), who sees irony as 'indirect' (i.e., inexplicit; cf. Utsumi, 2000) negation; related ideas are that of contrast (Colston, 2002) and inappropriateness (Attardo, 2000).

A very influential approach to irony is the mention theory (Sperber and Wilson, 1981), which claims that an utterance is ironical if it is recognized as the echoic mention of another utterance by a more or less clearly identified other speaker. Furthermore, the ironical statement must be critical of the echoed utterance (cf. Grice, 1989: 53–54). Similar theories based on the ideas of 'pretense' and 'reminder' have been presented as well. Criticism of the mention theory notes that not all irony seems to be interpretable as the echo of someone's words, or that if the definition of mention is allowed to encompass any possible mention it becomes vacuous (since any sentence is potentially the mention of another sentence). Furthermore, there exists an admittedly rarer, non-negative, praising irony, called *asteism* (Fontanier, 1968: 150). An example of *asteism* might be a colleague describing Chomsky's *Aspects of the theory of syntax* as a 'moderately influential' book in linguistics. Other approaches to irony include the 'tinge' theory, which sees irony as blending the two meanings (the stated and the implied ones) with the effect of attenuating the ironical one (Colston, 1997).

All the theories of irony mentioned so far share the idea that the processing of irony is a two-step process in which one sense (usually assumed to be the literal meaning) of the utterance is accessed and then a second sense of the utterance is discovered (usually under contextual pressure). Thus, for example, in a Gricean account of irony as implicature, the hearer of an utterance such as *That was smart* (uttered as a description of clumsy behavior, such as spilling one's wine upon someone's clothing) will first process the utterance as meaning literally roughly 'This behavior was consonant with how smart people behave' and then will discard this interpretation in favor of the implicature that the speaker means that the behavior was *not* consonant with how smart people behave. This account has been challenged recently by 'direct access' theories.

The direct access theories claim that the hearer does not process the literal meaning of an ironical

utterance first and only later accesses the figurative (ironical) meaning. Rather, they claim that the literal meaning is either not accessed at all or only later. Direct access interpretations of irony are squarely at odds with the traditional interpretation of irony as an implicature. Some results in psycholinguistics have been seen as supporting this view (Gibbs, 1994). The mention theory of irony was commonly interpreted as a direct access theory, but recent work (Yus, 2003) seems to indicate that it too can be interpreted as a two-step process. Other researchers (e.g., Dews and Winner, 1999) have presented contrasting views which support the two-step approach, although not always the claim that the literal meaning is processed first: claims that interpretations are accessed in order of saliency (Giora, 2003) or in parallel have been put forth.

Psycholinguistic studies of irony have focused on children's acquisition of irony (Winner, 1988), progressively lowering the age at which children understand irony to under ten years old; on the neurobiology of the processing of irony (McDonald, 2000), emphasizing the role of the right hemisphere alongside the left one (in which most language processing takes place); and on the order of activation of the various meanings in the ironical text. A significant issue is the degree and nature of the assumptions that the hearer and speaker must share for irony to be understood; this can be summed up as the 'theory of mind' that the speakers have. In particular, irony involves metarepresentations (Bara *et al.*, 1997, Curcó, 2000).

Considerable attention has been paid to the optional markers of irony, i.e., primarily intonational and kinesic indications of the speaker's ironical intent. While several phonological and other features have been considered 'markers' of irony, it appears that none of these features is exclusively a marker of irony. Reviews of markers include phonological (e.g., intonation), graphic (e.g., italics, punctuation), morphological (e.g., quotatives), kinesic (e.g., winking), and contextual clues (Haiman, 1998).

Recently, the social and situational context of irony as well as its pragmatic ends have begun being investigated in sociolinguistics and discourse/conversation analysis as well as in psycholinguistics. Work on the social functions of irony has found a broad range of functions, including in- and out-group definition, evaluation, aggression, politeness, verbal play, and many others (e.g., Clift, 1999; Anolli *et al.*, 2002; Gibbs and Colston, 2002; Kotthoff, 2003). It is likely that this list may be open-ended.

The relationship between irony and humor remains underexplored, despite their obvious connections, although some studies are beginning to address the interplay of irony and other forms of implicature, such

as indirectness, and metaphoricity. Finally, it is worth noting that dialogic approaches to language (e.g., Ducrot, 1984) see irony as a prime example of the co-presence of different 'voices' in the text, in ways that avoid the technical problems highlighted in the mention theories.

See also: Implicature; Relevance Theory; Speech Acts.

Bibliography

- Anolli L, Ciceri R & Riva G (eds.) (2002). *Say not to Say: New Perspectives on miscommunication*. Amsterdam: IOS Press.
- Anolli L, Infantino M G & Ciceri R (2002). "You're a real genius!": irony as a miscommunication design.' In Anolli, Ciceri & Riva (eds.). 135–157.
- Attardo S (2000). 'Irony as relevant inappropriateness.' *Journal of Pragmatics* 32(6), 793–826.
- Attardo S (2002). 'Humor and irony in interaction: from mode adoption to failure of detection.' In Anolli, Ciceri & Riva (eds.).
- Bara B, Maurizio T & Zettin M (1997). 'Neuropragmatics: neuropsychological constraints on formal theories of dialogue.' *Brain and Language* 59, 7–49.
- Booth W (1974). *A rhetoric of irony*. Chicago: University of Chicago Press.
- Clift R (1999). 'Irony in conversation.' *Language in Society* 28, 523–553.
- Colston H L (1997). 'Salting a wound or sugaring a pill: the pragmatic function of ironic criticism' *Discourse Processes* 23, 25–45.
- Colston H L (2002). 'Contrast and assimilation in verbal irony.' *Journal of Pragmatics* 34(2), 111–142.
- Curcó C (2000). 'Irony: negation, echo and metarepresentation.' *Lingua*, 110, 257–280.
- Dews S & Winner E (1999). 'Obligatory processing of literal and non-literal meanings in verbal irony.' *Journal of Pragmatics* 31(12), 1579–1599.
- Ducrot O (1984). *Le dire et le dit*. Paris: Editions de Minuit.
- Fontanier P (1968). *Les figures du discours*. Paris: Flammarion. Originally published as two volumes in 1821 and 1827.
- Gibbs R W (1994). *The poetics of mind: figurative thought, language, and understanding*. Cambridge/New York: Cambridge University Press.
- Gibbs R W & Colston H L (2002). 'The risks and rewards of ironic communication.' In Anolli, Ciceri & Riva (eds.). 181–194.
- Giora R (1995). 'On irony and negation.' *Discourse Processes* 19, 239–264.
- Giora R (2003). *On our mind*. Oxford: Oxford University Press.
- Haiman J (1998). *Talk is cheap: sarcasm, alienation, and the evolution of language*. Oxford/New York: Oxford University Press.
- Katz A N (ed.) (2000). 'The Uses and Processing of Irony and Sarcasm.' Special issue of *Metaphor and Symbol* 15(1/2).

-
- Kierkegaard S (1966). *The concept of irony, with constant reference to Socrates*. Capel L M (trans.). London: Collins.
- Kotthoff H (2003). 'Responding to irony in different contexts: on cognition in conversation.' *Journal of Pragmatics* 35(9), 1387–1411.
- Kreuz R J & Roberts R M (1995). 'Two cues for verbal irony: hyperbole and the ironic tone of voice.' *Metaphor and Symbolic Activity* 10(1), 21–31.
- McDonald S (2000). 'Neuropsychological studies of sarcasm.' *Metaphor and Symbol* 15(1/2), 85–98.
- Shelley C (2001). 'The bicoherence theory of situational irony.' *Cognitive Science* 25, 775–818.
- Sperber D & Wilson D (1981). 'Irony and the use mention distinction.' In Cole P (ed.) *Radical Pragmatics*. New York/London: Academic Press. 295–318.
- Toplak M & Katz A N (2000). 'On the uses of sarcastic irony.' *Journal of Pragmatics* 32(10), 1467–1488.
- Utsumi A (2000). 'Verbal irony as implicit display of ironic environment: distinguishing ironic utterances from nonirony.' *Journal of Pragmatics* 32(12), 1777–1806.
- Winner E (1988). *The point of words: children's understanding of metaphor and irony*. Cambridge, MA: Harvard University Press.
- Yus F (2003). 'Humor and the search for relevance.' *Journal of Pragmatics* 35(9), 1295–1331.
-

This page intentionally left blank



Language as an Object of Study

V Žegarac, University of Luton, Luton, UK

© 2006 Elsevier Ltd. All rights reserved.

Words are the leaves of the tree of language, of which,
if some fall away, a new succession takes their place.
– Field Marshall John French

What are the properties of language that make it similar to a tree? Could language exist without words in the way a tree may exist without leaves? Could words exist without the tree of language? One way to address these and other similar questions might be to make a list of the main properties of trees. The next step might be to look systematically for identical or similar properties in human languages. This would mean that our knowledge of botany would be guiding our scientific investigation of language. As fundamental differences between trees and languages are many and easily observable, it is evident that this research program does not hold enough promise to be worth pursuing. However, this general approach illustrates an important point about the methodology of scientific research in general and linguistic investigation of language in particular: research into one poorly understood object of study proceeds on the assumption that it shares its essential properties with some other (type of) object which is already better understood. In fact, some major landmarks in the development of modern linguistics are based on four conceptions of language as an object of study: language as a social fact, language as behavior, language as a mental organ, and language as an abstract object.

Language as a Social Fact

The view of language as a social fact, as suggested by Ferdinand de Saussure (1916), marked a shift away from the predominantly historical interest in language toward the study of language as a system at the current stage of its development. Saussure believed that important insights into language could be gained if it was investigated from the point of view of its users, who typically do not know anything about its historical

development. Since language use reflects the structure of language as a system of units at a particular stage of its development, language should be studied from the synchronic, rather than the diachronic, perspective. Saussure compared a given stage in the development of a language with the configuration of pieces on the chess board at a given stage of a game of chess. Just as the value of each piece is determined by its position relative to all the other pieces on the board (at a given stage of the game), a particular language at a given stage of development is a system of language units, or language signs (Saussure's term is *signe linguistique*), whose values are determined by their position in relation to all the other signs in the language system at that particular stage. And just as any change in the position of a given chess piece alters the values of all the other pieces and affects the system as a whole, the change that directly affects one language sign (potentially) affects the language system as a whole. For example, although the Serbian word 'jeftin' has the same meaning as the English word 'cheap,' it does not have the same value in the language system, because, in English, 'cheap' contrasts with 'inexpensive,' whereas in Serbian the same contrast is not lexicalized, so that it is either ignored, with 'jeftin' covering both meanings, or it is conveyed in some more elaborate way. The value of the Serbian word 'jeftin,' – its position in the language system – would change significantly if the equivalent of the English 'inexpensive' were to become lexicalized in Serbian.

In Saussure's view, language is a social fact akin to other social institutions, such as the legal system or the dress code. Although people's use of language reflects its systematic character, the structure of language as a system is not directly observable or consciously represented by its users. Saussure posited a distinction between observable linguistic behavior, what he called *parole* (speaking), and the underlying language system, what he called *langue* (language). In his view, the main task of linguistics was to discover *langue* – the language system – by investigating *parole* – language in use. Saussure's view of languages as a social fact is consistent with Émile Durkheim's

sociological theory (popular at the time when Saussure was developing his ideas on language), in which social facts are characterized as representations in the ‘collective mind,’ rather than as material or psychological entities.

Saussure’s concept of *langue* as a social fact, and as the proper object of linguistic investigation, had an unfortunate consequence. If *langue* is a system of relations between language signs, which are both external to its users and not directly observable in the manifestations of *langue* in use (because they are part of the ‘collective mind’), there is no place for the category of sentence in linguistic analysis. In other words, it follows from Saussure’s position that sentences are an aspect of *parole* rather than *langue*. Therefore, they are not a part of the systematic character of a language. So, although we readily produce and comprehend utterances of sentences that we have never produced or heard before and, although we assign them meanings in a systematic way, Saussure’s view of language as a social fact did not adequately accommodate this important aspect of language.

Language as Behavior

The view of language as behavior emerged in American descriptivist linguistics under the influence of positivism (in philosophy) and behaviorism (in psychology). An important methodological stance of logical positivism is that scientific research must be based either on statements that are tautologically true, such as: either P or not P, or they must be based on direct observation. The immediate consequence of this view for the scientific study of language was that all analyses and generalizations should be based strictly on the description of observable linguistic behavior. The main methodological stance of behaviorist psychology was that scientific theories could be refuted only by observable phenomena. Hence the descriptivists’ view that insights into language should be firmly grounded in observable data – the products of linguistic behavior – and that they should not be informed by the researcher’s introspection or the native speaker’s intuition. This meant that more attention was paid to particular languages than to the universal properties of human language, as we can only have direct data from particular languages. However, although descriptivist linguists initially thought of human languages as infinitely diverse, by the 1950s their interest had shifted towards the possibility of devising a discovery procedure, a mechanism that would take a corpus of utterances as input and yield a grammar as output (Harris, 1951).

The most important theoretical tenet of behaviorism was that all human behavior could be explained in terms of stimuli and responses to stimuli without reference to mental structures and processes. This stance had two major implications for the study of language: first, language was characterized as “the totality of utterances that can be made in a speech community” (Bloomfield, 1928). Second, since meaning is not directly observable and it could not be studied without reference to intuitive, introspective judgments, it was considered not to be the proper object of linguistics. Though to a lesser extent than Saussure’s conception of language as a social fact, the view of language as the product of behavior (especially in the early stages) also tended to focus on the phonological and morphological analysis of language, rather than on the less directly observable structural relations between words within sentences.

Language as a Mental Organ

The view of language as a mental organ is central to the most influential modern linguistic theory, the generative approach to language, developed by Noam Chomsky since the 1950s and his many followers (see Chomsky, 1986, 1988, 2000). Although the system of ideas that characterizes present-day generative linguistics is very complex, the basic assumptions on the nature of language are clearly established and they appear to be fundamentally commonsensical.

The idea that language is a mental organ is best illustrated by an analogy with other biologically specified properties of human beings. The development of any organism is partly determined by its genetic makeup and partly by environmental factors. For example, the onset of puberty and the height of a person are genetically determined, but they are also affected to some limited extent by external factors, such as nutrition. Given the diversity of environmental circumstances in which people grow up, and the uniformity in their physical appearance, there can be hardly any doubt as to the primacy of genetic makeup over the environment. The same observation carries over to our mental capacities. For example, numerical ability is unique to humans, who develop it in a uniform way regardless of the differences in their sociocultural and physical environments. Similarly, the design features of language, such as recursion, arbitrariness, and duality of patterning, cannot be explained as the result of environmental factors. Therefore, we should assume that both our ability to count and our linguistic ability are genetically specified. In other words, just as arms and legs are physical organs, language is a mental organ. The interaction

with the environment triggers the development of language, but the environmental inputs are simply not sufficiently structured for it to be plausible to assume that language development involves the construction of more and more detailed mirror images of the external world. Other comparable capacities would include our musical ability and the ability to construct scientific theories.

If language is a mental organ, then linguistics should be seen as a discipline of psychology, ultimately biology. It should aim to discover the abstract properties of language that make it possible to answer the following questions:

1. What do we know when we know language?
2. How is knowledge of language acquired?
3. How is this knowledge of language put to use?

In this view, the knowledge of language is characterized as a (generative) grammar: a finite system of interacting rules and principles that specify (more technically, generate) an infinite number of expressions, each of which is a structural configuration (e.g., of words into phrases and sentences) associated with a phonetic form and meaning. At the core of our ability to acquire language lie some universal innate properties of grammar, which guide the child's language development. The innate properties of grammar constrain the range of possible grammars of language and enable the child to find out which of the possible grammars is best evidenced by the linguistic behavior of those around him. This genetic endowment enables the child to go beyond the evidence presented by the data in determining the grammar. For instance, a child acquiring English figures out that 'John is too stubborn to talk to' means, roughly, 'John is so stubborn that nobody can talk to him' without explicit instruction and without being distracted by the possible analogy with utterances like 'John is too stubborn to talk to Bill,' which would be misleading. Therefore, evidence from language acquisition supports the view that the child constructs the grammar on the basis of innate mental structures that heavily constrain the range of available choices at any given stage of his linguistic development. The question of how knowledge of language is put to use is investigated by designing 'performance models' based on the mental grammar.

The research program of Chomskyan generative grammar differs in two important respects from those associated with the view of language as a social fact and the view of language as behavior. First, there is a shift of emphasis from the study of individual languages to the study of Universal Grammar (UG), where the latter is another name for language as a mental organ (in terms of the properties it identifies as

the set of grammars that a human being can learn under the conditions in which language acquisition normally takes place). Second, language as an object of study is an element of the mind, which Chomsky calls Internalized language (I-language). The approaches that consider language to be a social fact or behavior and the products of behavior are concerned with the study of language as external to the minds of its speakers. In Chomsky's terms, they are concerned with the study of Externalized language (E-language). Chomsky argues that E-languages (which correspond closely to what people normally mean when they speak about English, French, German, Chinese, Swahili, etc.) are epiphenomenal. In other words, they are complex constructs resulting from the interaction of various sociopolitical and historical factors. Therefore, understood in this way, a particular language like English or French is not a proper object of scientific investigation. Although people ordinarily speak about particular languages as individual objects that exist independently of individual speakers (in utterances like 'French is the language of diplomacy.'). Chomsky insists that this shift of focus from E-language to I-language is a shift towards the commonsense view of language, because when we say that a person knows a language we usually mean that they know how to relate sound patterns with meanings in a particular systematic way.

Language as an Abstract Object

In contrast to Chomsky's mentalist conception of language, Katz (1981) presents a detailed argument to support the view of language as an abstract object. The distinctive characteristics of abstract objects are (a) that their existence is independent of the existence of the mind, and (b) that they do not occupy a position in space and time. Presumably, natural numbers, which underlie the laws of nature, would exist even if human or other intelligence did not, and they are not located at particular places and times. In the light of these observations a natural language, such as English or Japanese, might seem a poor candidate for the category of abstract object. However, Katz (1981) argues that the task of writing the grammar of a language is too stringently constrained on Chomsky's requirement "that a grammar represent a language only in the particular form that knowledge of the language takes when such knowledge is realized in the human mind or brain" (Katz, 1981: 92). Arguably, this requirement has some unwelcome implications for linguistic analysis. For example, let us say that two grammars of a given language (G1 and G2) have been written. G1 reflects the speaker-hearer's

knowledge better than G2, but G2 is simpler and more elegant than G1. G1 and G2 are equivalent in all other respects. The question is: 'Is G1 to be preferred to G2, or is G2 to be preferred to G1?' In Katz's view it would be a mistake to reject G2 just because it is less psychologically plausible than G1. He points out the importance of the distinction between the knowledge of something and the object of knowledge itself. For example, natural numbers are distinct from the different ways in which they are represented in various calculators. By the same token, the speaker's knowledge of language is distinct from language itself. Chomsky (1986: 49) rejects this view with the following analogy between language and the heart as biological organs: the heart may be simply and elegantly described as a pump, disregarding what actually makes it beat as a part of the organism, but it does not follow that such a description would be superior to one which looks at the actual anatomy and physiology of the human heart. Therefore, if G2 is simpler and more elegant than G1, but G1 is more plausible as a psychologically real description of the speaker-hearer's knowledge of their language, then it is obvious that G2 should not be preferred to G1.

Katz also challenges Chomsky's view that individual languages are not proper objects of scientific study because they are sociopolitical constructs. In his view, this is "like claiming that the concept of number is not an object of mathematics but a sociopolitical one." (1981: 79–80). Chomsky (1986: 47) rejects this objection out of hand by describing it as a "curious conclusion" (for other criticisms and a defense against them, see Katz, 1985). Although some criticisms of Katz's view of language seem very convincing, a reasonably strong case can be made in support of the conception of language as an abstract object. On the one hand, the idea that language can be seen and studied scientifically as an abstract object does not entail that it is not also a psychological object. Hence, it is far from obvious that the only interesting and useful scientific approach to the study of language is psychological (biological). In other words, the conceptions of language as a mental organ and language as an abstract object might both provide the basis for credible scientific research. Moreover, the scientific study of various physical organs, such as the human heart, can easily draw on readily available evidence of their structure. By comparison, the evidence for the analysis and description of language as a mental organ is rather scant. Thus, if an organ, e.g., the human heart, were not available for relatively direct observation, then it would be perfectly sensible to investigate it by trying to figure

out the best abstract model in the hope that it would also turn out to be the one that corresponds most closely to the real thing. In fact, much work within Chomsky's generative approach in linguistics seems to proceed in this way. On the other hand, Katz (1985: 199–200) argues that theories of language grounded in individual psychology are not abstract enough to provide plausible explanations of language structure. His discussion of analytic sentences is particularly convincing. Analytic sentences, such as 'Nightmares are dreams.' and 'Flawed gems are imperfect.' are necessarily true in virtue of the semantic level of grammars. The Chomskyan approach, which characterizes the grammar in psychological terms, cannot provide an adequate account of necessary truths. 'Nightmares are dreams.' and 'Flawed gems are imperfect.' are true sentences, regardless of our psychological (biological) makeup, but, in the Chomskyan approach, these sentences have to be analyzed as necessarily true, because humans are psychologically (biologically) designed so that we cannot think of them as anything other than true statements. While this and related issues remain open for debate, there can be no doubt that Chomsky's psychological approach to language remains the most influential perspective from which it is studied within linguistics.

Bibliography

- Bloomfield L (1928). 'A set of postulates for the science of language.' *Language* 2. In Joos M (ed.) (1957) *Readings in linguistics*. Washington, DC: American Council of Learned Sciences, 1957.
- Bloomfield L (1933). *Language*. New York: Holt.
- Chomsky N (1986). *Knowledge of language: its nature, origin and use*. New York: Praeger.
- Chomsky N (1988). *Language and problems of knowledge: the Managua lectures*. Cambridge, MA: MIT Press.
- Chomsky N (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- de Boysson-Bardies B (1999). *How language comes to children*. Cambridge, MA: MIT Press.
- de Saussure F (1916). *Cours de linguistique générale*. Bally C & Sechaye A (eds.). Paris: Payot, 1968.
- Harris Z S (1951). *Structural linguistics*. Chicago: The University of Chicago Press.
- Katz J (1981). *Language and other abstract objects*. Oxford: Basil Blackwell.
- Katz J (1985). 'An outline of Platonist grammar.' In Katz J (ed.) *The philosophy of linguistics*. Oxford: Oxford University Press.
- Pinker S (1994). *The language instinct: the new science of language and mind*. London: Penguin Press.
- Sampson G (1980). *Schools of linguistics: competition and evolution*. London: Hutchinson.

Lexical Conceptual Structure

J S Jun, Hankuk University of Foreign Studies,
Seoul, Korea

© 2006 Elsevier Ltd. All rights reserved.

Introduction

The lexical conceptual structure (LCS) or simply the conceptual structure (CS) is an autonomous level of grammar in conceptual semantics (Jackendoff, 1983, 1990, 1997, 2002), in which the semantic interpretation of a linguistic expression is explicitly represented. Jackendoff's (1983) original conception is to posit a level of mental representation in which *thought* is couched (cf. the *language of thought* in Fodor, 1975). CS is a relay station between language and peripheral systems such as vision, hearing, smell, taste, kinesthesia, etc. Without this level, we would have difficulty in describing what we see and hear.

There are two ways to view CS in formalizing a linguistic theory. One is to view CS as a nonlinguistic system that serves as an interface between meaning and nonlinguistic modalities. Then, we need another level of representation for meaning (cf. Chomsky's [1981, 1995] LF); and CS is related to the linguistic meaning by pragmatics as shown in Figure 1. This is the view of Katz and Fodor (1963), Jackendoff (1972), Katz (1980), and Bierwisch and Schreuder (1992). The alternative conception is to view CS as the semantic structure. The linguistic meaning as well as nonlinguistic information compatible with sensory and motor inputs is directly represented in CS. CS is related with other linguistic levels such as syntax and phonology by correspondence rules, and therefore CS is part of the lexical information (*hence called*

LCS) as shown in Figure 2. This is the current view of conceptual semantics.

One argument that supports the latter view comes from generic judgment sentences. In the standard view of linguistic meaning, judgments of superordination, subordination, synonymy, entailment, etc., are *linguistic*. We judge that 'bird' and 'chicken' make a superordinate-subordinate pair; that in some dialects 'cellar' and 'basement' are synonymous; and that 'Max is a chicken' entails 'Max is a bird.' Linguistic judgments of this sort are formalized in theories such as meaning postulates (Fodor, 1975) and semantic networks (Collins and Quillian, 1969).

Jackendoff (1983) points out one problem in formalizing these judgments from a purely linguistic perspective: judgments of superordination and subordination, for instance, are directly related to judgments of generic categorization sentences such as 'A chicken is a bird.' The judgment about generic categorization is, however, not entirely linguistic or semantic, in that it behaves creatively enough to include ambiguous cases such as (1) below.

- (1a) A piano is a percussion instrument.
 - (1b) An australopithecine was a human.
 - (1c) Washoe (the chimp)'s sign system is a language.
 - (1d) An abortion is a murder.
- (Jackendoff, 1983: 102)

We make generic categorization judgments about (1) not on the basis of meaning postulates or semantic networks but on the basis of our factual, often political, world knowledge. For instance, our judgment about (1d) is influenced by our political position, religion, and knowledge about biology. This is analogous to Labov's (1973) dubious 'cup-bowl'

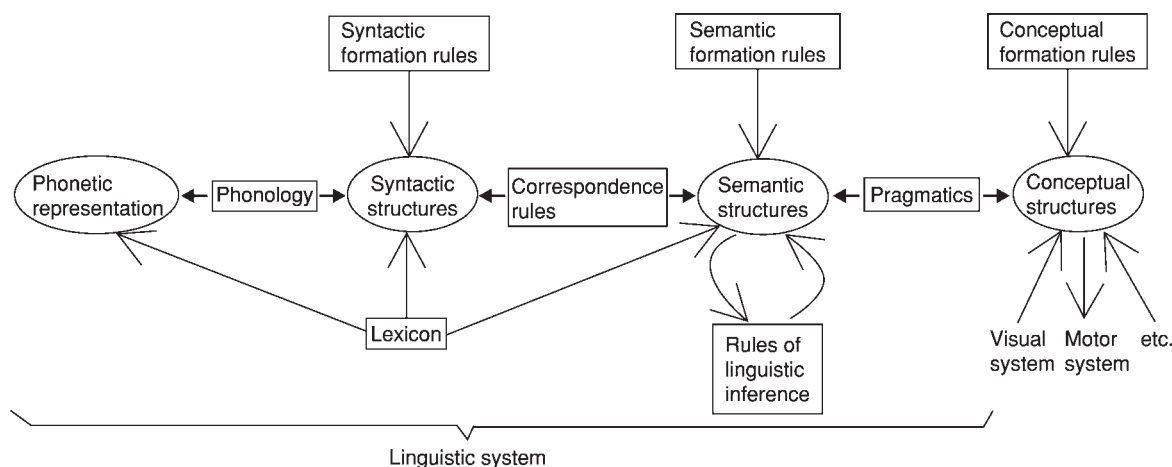


Figure 1 CS as a nonlinguistic system (adapted from Jackendoff R (1983). *Semantics and cognition*. Cambridge, MA: MIT Press, 20, with permission).

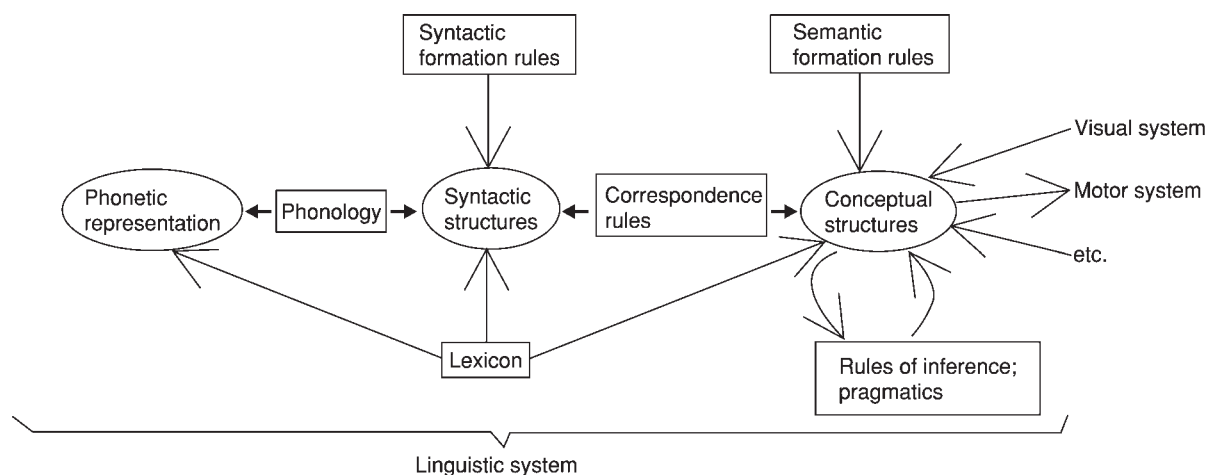


Figure 2 CS as part of the linguistic system (adapted from Jackendoff R (1983). *Semantics and cognition*. Cambridge, MA: MIT Press, 21, with permission).

judgment, which obviously resorts to nonlinguistic encyclopedic knowledge as well as the linguistic type system. CS is, by definition, the level that represents encyclopedic knowledge as part of our thought. Hence, we should refer to CS to make generic categorization judgments about (1).

Jackendoff's (1983) puzzle is summarized, as follows. We make judgments of semantic properties such as superordination and subordination at the level of semantic structure. We make generic categorization judgments at the level of CS as shown by (1). If the semantic structure were separated from CS, we would fail to catch the obvious generalization between the superordinate-subordinate judgment and the generic categorization judgment. If, by contrast, CS were the semantic structure, we would have no trouble in accounting for the intuitive identity between the two judgments. Therefore, CS is the semantic structure. For more arguments to support the view that CS is the semantic structure, see Jackendoff (1983: Ch. 6).

Overview of Conceptual Semantics

Autonomy of Semantics

A central assumption in conceptual semantics is the autonomy of semantics. In Chomsky's view of language, syntax makes an autonomous level of grammar, whereas phonology and semantics merely serve as interpretive components (PF and LF). Jackendoff (1997) criticizes this view as *syntactocentric*, and provides convincing arguments to support his thesis that phonology and semantics as well as syntax make autonomous levels of grammar.

We find numerous pieces of evidence for the autonomy of semantics in the literature of both psycholinguistics and theoretical linguistics. Zurif and Blumstein's (1978) pioneering work shows that Wernicke's area is the center of semantic knowledge in the brain in comparison with Zurif, Caramazza and Myerson's (1972) previous finding that Broca's area is the center of syntactic knowledge. Swinney's (1979) classical work on lexical semantic priming shows that lexical semantics is independent of the grammatical contexts like the movement chain in a sentence. Piñango, Zurif, and Jackendoff (1999) report more workload for the online processing of aspectual coercion sentences (e.g., *John jumped for two hours*) than for the processing of syntactically equivalent noncoerced sentences (e.g., *John jumped from the stage*).

Semantic categories are not in one-to-one correspondence with syntactic categories. For instance, all physical object concepts correspond to nouns, but not all nouns express physical object concepts; e.g., *earthquake* and *concert* express event concepts. All verbs express event/state concepts, but not all event/state concepts are expressed by verbs; e.g., *earthquake* and *concert* are nouns.

Contrary to Chomsky's (1981) theta criterion, we have plenty of data that shows mismatch between syntactic functions and thematic roles. For instance, the semantic interpretation of *buy* necessarily encodes both the transfer of money from the buyer to the seller and the transfer of the purchased entity from the seller to the buyer. Among the three semantic arguments, i.e., the buyer, the seller, and the purchased object, only the buyer and the purchased entity are syntactic arguments (e.g., *John bought the book*).

The seller is syntactically expressed as an adjunct (e.g., *John bought the book from Jill*). Moreover, the buyer plays the source role of money and the target role of the purchased entity simultaneously; the seller plays the source role of the purchased entity and the target role of money simultaneously. In short, the buyer and the seller have multiple theta roles even though each of them corresponds to one and only one syntactic entity.

A simple semantic distinction often corresponds to many syntactic devices. For instance, telicity is expressed by such various syntactic devices as choice of verb (2a), choice of preposition (2b), choice of adverbial (2c), choice of determiner in the subject NP (2d) and in the object NP (2e), and choice of prepositional object (2f) (Jackendoff, 1997: 35).

- | | |
|---|----------|
| (2a) John <i>destroyed</i> the cart (in/*for an hour). | → Telic |
| John <i>pushed</i> the cart (for/*in an hour). | → Atelic |
| (2b) John ran <i>to</i> the station (in/*for an hour). | → Telic |
| John ran <i>toward</i> the station (for/*in an hour). | → Atelic |
| (2c) The light flashed <i>once</i> (in/*for an hour). | → Telic |
| The light flashed <i>constantly</i> (for/*in an hour). | → Atelic |
| (2d) <i>Four people</i> died (in/*for two days). | → Telic |
| <i>People</i> died (for/*in two days) | → Atelic |
| (2e) John ate <i>lots of peanuts</i> (in/*for an hour) | → Telic |
| John ate <i>peanuts</i> (for/*in an hour). | → Atelic |
| (2f) John crashed into <i>three walls</i> (in/*for an hour) | → Telic |
| John crashed into <i>walls</i> (for/*in an hour) | → Atelic |

To sum up, the mapping between syntax and semantics is not one-to-one; rather, it is one-to-many, many-to-one, or at best many-to-many. The mapping problem is not easy to explain in the syntactocentric

architecture of language. The overall difficulty in treating semantics merely as an interpretive component of grammar along with a similar difficulty treating phonology as an interpretive component (cf. Jackendoff, 1997: Ch. 2) leads Jackendoff to propose a tripartite architecture of language, in which phonology, syntax, and semantics are all independent levels of grammar licensed by phonological formation rules, syntactic formation rules, and conceptual/semantic formation rules respectively, and interfaced by correspondence rules between each pair of modules, as shown in Figure 3.

Lexical Conceptual Structure

Conceptual semantics assumes striking similarities for the organization of CS with the structural organization of syntax. As syntax makes use of syntactic categories, namely syntactic parts of speech like nouns, adjectives, prepositions, verbs, etc., semantics makes use of semantic categories or semantic parts of speech such as Thing, Property, Place, Path, Event, State, etc. As syntactic categories are motivated by each category member's behavioral properties in syntax, semantic or ontological categories are motivated by each category member's behavioral properties in meaning.

Syntactic categories are combined by syntactic phrase-structure structure rules into larger syntactic expressions; likewise, semantic categories are combined by semantic phrase-structure rules into larger semantic expressions. The syntactic representation is

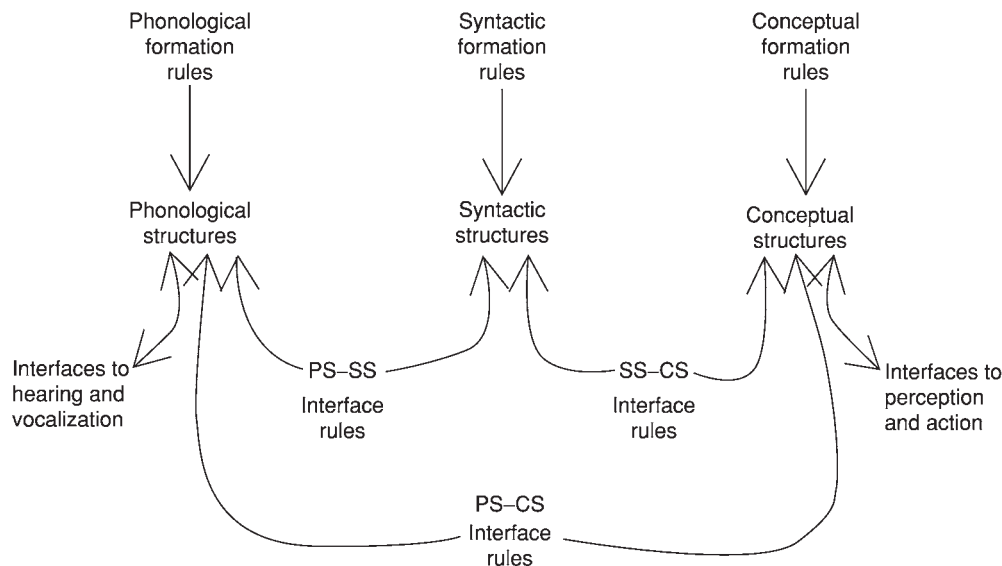


Figure 3 The tripartite parallel architecture (reproduced from Jackendoff R (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press).

structurally organized, so we can define dominance or government relations among syntactic constituents; likewise, the semantic representation is structurally organized, so we can define grammatically significant hierarchical relations among semantic constituents. Various syntactic phrase-structure rules can be generalized into a rule schema called X-bar syntax (Jackendoff, 1977); likewise, various semantic phrase-structure rules can be generalized into a rule schema called X-bar semantics (Jackendoff, 1987b).

Ontological Categories Ontological categories are first motivated by our cognitive layouts. To mention some from the vast psychology literature, Piaget's developmental theory of *object permanence* shows that infants must recognize objects as a whole, and develop a sense of permanent existence of the objects in question when they are not visible to the infants. Researchers in language acquisition have identified many innate constraints on language learning like reference principle, object bias, whole object principle, shape bias, and so on (cf. Berko Gleason, 1997). For instance, children rely on the assumption that words refer to objects, actions, and attributes in the environments by reference principle. Wertheimer's (1912) classical experiment on *apparent movement* reveals that humans are equipped with an innate tendency to perceive the change of location as movement from one position to the other; the apparent movement experiment renders support for the expansion of the event category into function argument structures like $[_{Event} GO ([_{Thing}], [_{Path}])]$.

Ontological categories also have numerous linguistic motivations. Pragmatic anaphora (exophora) provides one such motivation. In order to understand the sentence in (3), the hearer might have to pick out the referent of *that* among several entities in the visual field. If the hearer did not have object concepts to organize the visible entities, (s)he could not pick out the proper referent of the pragmatic anaphora *that*. The object concept involved in the semantic interpretation of (3) motivates the ontological category Thing.

(3) I bought *that* last night.

The category Thing proves useful in interpreting many other grammatical structures. It provides the basis of interpreting the Wh-variable in (4a); it supports the notion of identity in *the same* construction in (4b); and it supports the notion of quantification as shown in (4c).

- (4a) *What* did you buy last night?
 (4b) John bought *the same thing* as Jill.

(4c) John bought *something/everything that Jack bought*.

Likewise, we find different sorts of pragmatic anaphora that motivate ontological categories like Place (5a), Direction (5b), Action (5c), Event (5d), Manner (5e), and Amount (5f).

- (5a) Your book was *here/there*.
 (5b) They went *there* yesterday.
 (5c) Can he do *this/that*?
 (5d) *It* happened this morning.
 (5e) Bill shuffled a deck of cards *this* way.
 (5f) The man I met yesterday was *this* tall.

These ontological categories provide innate bases for interpreting Wh-variables, the identity construction, and the quantification, as shown in (6)–(8).

- (6a) *Where* was my book?
 (6b) *Where* did they go yesterday?
 (6c) *What* can he do?
 (6d) *What* happened this morning?
 (6e) *How* did Bill shuffle a deck of cards?
 (6f) *How* tall was the man you met yesterday?
 (7a) John put the book on *the same place* as Bill.
 (7b) John went *the same way* as Bill.
 (7c) John did *the same thing* as Bill.
 (7d) *The same thing* happened yesterday as happened this morning.
 (7e) John shuffled a deck of cards *the same way* as Bill.
 (7f) John is *as tall as* the man I met yesterday.
 (8a) John put the book at *some place that Bill put it*.
 (8b) John went *somewhere that Bill went*.
 (8c) John did *something Bill did*.
 (8d) *Something that happened this morning* will happen again.
 (8e) John will shuffle cards in *some way that Bill did*.
 (8f) (no parallel for amounts)

For more about justifying ontological categories, see Jackendoff (1983: Ch. 3).

Conceptual Formation Rules Basic ontological categories are expanded into more complex expressions using function-argument structural descriptions. (9) shows such expansions of some ontological categories.

- (9a) EVENT \rightarrow $[_{Event} GO (THING, PATH)]$
 (9b) EVENT \rightarrow $[_{Event} STAY (THING, PLACE)]$
 (9c) EVENT \rightarrow $[_{Event} CAUSE (THING \text{ or } EVENT, EVENT)]$
 (9d) EVENT \rightarrow $[_{Event} INCH (STATE)]$
 (9e) STATE \rightarrow $[_{State} BE (THING, PLACE)]$
 (9f) PLACE \rightarrow $[_{Place} PLACE-FUNCTION (THING)]$
 (9g) PATH \rightarrow $[_{Path} PATH-FUNCTION (THING)]$

The function-argument expansion is exactly parallel with rewriting rules in syntax (e.g., $S \rightarrow NP VP$; $NP \rightarrow Det (AP)^* N$; $VP \rightarrow V NP PP$), and hence can be regarded as semantic phrase-structure rules. The semantic phrase-structure rules in (9) allow recursion such as syntactic phrase-structure rules: an Event category can be embedded in another Event category as shown in (9c). We also can define hierarchical relations among conceptual categories in terms of the depth of embedding as we define syntactic dominance or government in terms of the depth of embedding in syntactic structures. The depth of embedding in CS plays a significant role in explaining such various grammatical phenomena as subject selection, case, binding, control, etc. See Culicover and Jackendoff (2005) for more about these issues.

Place functions in (9f) may include IN, ON, TOP-OF, BOTTOM-OF, etc. Path functions in (9g) may include TO, FROM, TOWARD, VIA, etc. Conceptual semantics is a parsimonious theory, in that it makes use of only a handful of functions as conceptual primitives. All functions should be motivated on strict empirical grounds. This is exactly parallel with using only a handful of syntactic categories motivated on strict empirical grounds. Syntactic phrase-structure rules do not refer to unlimited number of syntactic categories. Syntactic categories such as noun, adjective, preposition, verb, etc. are syntactic primitives, and they are motivated by each category member's behavioral properties in syntax. Likewise, semantic phrase-structure rules refer to a restricted set of semantic or conceptual primitives that are empirically motivated by general properties of meaning.

Functions such as GO, BE, and STAY are empirically motivated in various semantic fields. They are the bases for interpreting spatial sentences in (10).

- (10a) GO: The train traveled from Boston to Chicago.
- (10b) BE: The statue stands on Cambridge common.
- (10c) STAY: John remained in China.

These functions also support the interpretation of possession sentences in (11).

- (11a) GO: John gave the book to Bill.
- (11b) BE: John had no money.
- (11c) STAY: The library kept several volumes of the Korean medieval literature.

Interpreting ascription sentences also require GO, BE, and STAY, as shown in (12).

- (12a) GO: The light turned from yellow to red.
- (12b) BE: The stew seemed distasteful.
- (12c) STAY: The aluminum stayed hard.

One interesting consequence of having GO, BE, and STAY in both spatial and nonspatial semantic fields is that we can explain how we use the same verb for different semantic fields.

- (13a) The professor turned into a driveway. (Spatial)
- (13b) The professor turned into a pumpkin. (Ascription)
- (14a) The bus goes to Paris. (Spatial)
- (14b) The inheritance went to Bill. (Possession)
- (15a) John is in China. (Spatial)
- (15b) John is a doctor. (Ascription)
- (16a) John kept the CD in his pocket. (Spatial)
- (16b) John kept the CD. (Possession)
- (17a) The professor remained in the driveway. (Spatial)
- (17b) The professor remained a pumpkin. (Ascription)

In (13), the verb *turn* is used in both spatial and ascription sentences with the GO meaning. How do we use the same verb for two different semantic fields? Do we have to assume two different lexical entries for *turn*? Conceptual semantics does not pay anything to explain this puzzle. We do not need two different lexical entries for *turn* to explain the spatial and ascription meanings. We just posit the event function GO for the lexical semantic description or LCS for *turn* in (13). Both spatial and ascription meanings follow from the LCS for *turn*, since the function GO is in principle motivated by both spatial and ascription sentences. We can provide similar accounts for all the data in (14)–(17). For more about the general overview of conceptual semantics, see Jackendoff (1983, 1987a, 1990, 2002).

X-bar Semantics

Generative linguists in the 1950s and 1960s succeeded in showing the systematic nature of language with a handful of syntactic phrase-structure rules. But they were not sure how the phrase-structure rules got into language learners' minds within a relatively short period of time; it was a learnability problem. X-bar syntax (Chomsky, 1970; Jackendoff, 1977) opened a doorway to the puzzle. Children do not have to be born with dozens of syntactic categories; children are born with one syntactic category, namely, category X. Children do not have to learn dozens of totally unrelated syntactic phrase-structure rules separately; all seemingly different syntactic phrase-structure rules share a fundamental pattern, namely, X-bar syntax.

Jackendoff (1987b, 1990), who was a central figure in developing X-bar syntax in the 1970s, has completed his X-bar theory by proposing X-bar semantics. We have so far observed that CS is exactly parallel with the syntactic structure. Conceptual categories are structurally organized into CS by virtue of semantic phrase-structure rules, as syntactic categories are structurally organized into syntactic structure by virtue of syntactic phrase structure rules. (18) is the basic formation of X-bar syntax.

- (18a) $XP \rightarrow \text{Spec } X'$
 (18b) $X' \rightarrow X \text{ Comp}$
 (18c) $X \rightarrow [\pm N, \pm V]$

Now that CS has all parallel properties with the syntactic structure, all semantic phrase-structure rules are generalized into X-bar semantics along the same line with X-bar syntax as shown in (19).

$$(19) [\text{Entity}] \rightarrow \begin{bmatrix} \text{Event Thing Place} \dots \\ \text{Token Type} \\ F(\langle \text{Entity}_1, \langle \text{Entity}_2, \langle \text{Entity}_3 \rangle \rangle \rangle) \end{bmatrix}$$

(19) provides not only the function-argument structural generalization for all the semantic phrase-structure rules but also shows how major syntactic constituents correspond to major conceptual categories. That is, the linking between syntax and semantics can be formalized as (20) and (21).

$$(20) XP \text{ corresponds to } [\text{Entity}]$$

$$(21) \begin{bmatrix} X_0 \\ \text{---} \langle YP \langle ZP \rangle \rangle \end{bmatrix} \underset{\text{to}}{\text{corresponds to}} \begin{bmatrix} \text{Entity} \\ F(E_1, \langle E_2, \langle E_3 \rangle \rangle) \end{bmatrix}$$

where YP corresponds to E_2 , ZP corresponds to E_3 , and the subject (if there is one) corresponds to E_1 .

To sum up, the obvious similarity between (18) and (19) enables us to account for the tedious linking problem without any extra cost.

General Constraints on Semantic Theories

Jackendoff (1983) suggests six general requirements that any semantic theory should fulfill: expressiveness, compositionality, universality, semantic properties, the grammatical constraint, and the cognitive constraint.

First, a semantic theory must be observationally adequate; it must be expressive enough to describe most, if not all, semantic distinctions in a natural language. Conceptual semantics has expressive power, in that most semantic distinctions in a natural language can be represented by CS with a handful of conceptual categories plus conceptual formation rules. What is better is that the expressive power has improved since the original conception of the theory. For instance, Jackendoff (1990: Ch. 7) introduced the action tier

into the theory to represent the actor/patient relation aside from motion and location. In (22a), *John* is the source of the *ball* and the actor of the throwing event simultaneously; the *ball* is a moving object, the theme, and an affected entity, the patient, simultaneously. It is quite common for one syntactic entity to bear double theta roles contra Chomsky's (1981) theta criterion; conceptual semantics captures this by representing the motion/location event in the thematic tier (22b), and the actor/patient relation in the action tier (22c).

- (22a) John threw the ball.
 Source Goal
 Actor Patient
 (22b) $[\text{Event CAUSE}([\text{JOHN}], [\text{Event GO}([\text{BALL}],$
 $[\text{Path TO}([\dots])]])]$
 (22c) $[\text{AFF}([\text{JOHN}], [\text{BALL}])]$

The action tier not only explains the fine semantic distinction in language but also plays a central role in such grammatical phenomena as linking and case. Besides the action tier, Jackendoff (1991) introduced an elaborate feature system into CS to account for the semantics of parts and boundaries; Csuri (1996) introduced the referential tier into CS that describes the definiteness of expressions; Jackendoff (2002) introduced the lambda extraction and the topic/focus tier into CS. All these and many other innovations make the theory expressive enough to account for significant portion of natural language semantics.

The second constraint on a semantic theory is compositionality: an adequate semantic theory must show how the meanings of parts are composed into the meaning of a larger expression. Conceptual semantics is compositional, in that it shows how combinatorial rules of grammar compose the meanings of ontological categories into the CS of a larger expression.

The third requirement is universality: an adequate semantic theory must provide cross-linguistically relevant semantic descriptions. Conceptual semantics is not a theory of meaning for any particular language. It is a universal theory of meaning; numerous cross-linguistic studies have been conducted with the conceptual semantic formalism. See Jun (2003), for instance, for a review of many conceptual semantic studies on the argument linking and case in languages such as Korean, Japanese, Hindi, Urdu, English, Old English, French, etc.

The fourth requirement is semantic properties: an adequate semantic theory should be able to explain many semantic properties of language like synonymy, anomaly, presupposition, and so on. That is, any semantic theory must explicate the valid inference of expressions. CS provides a direct solution to this problem in many ways. The type/token distinction is directly expressed in CS, and explains most semantic

distinctions made by the semantic type system. By decomposing verbs such as *kill* into [CAUSE ([THING], [NOT-ALIVE ([THING])])], conceptual semantics explains how *John killed Bill* entails *Bill is dead*. For more about semantic properties, see Jackendoff (1983, 1990, 2002).

The fifth requirement is the grammatical constraint: if other things were equal, a semantic theory that explains otherwise arbitrary generalizations about the lexicon and the syntax would be highly preferable. Conceptual semantics is a theory of meaning that shows how a handful of conceptual primitives organize the vast domain of lexical semantics. Conceptual semantics also explains how semantic entities are mapped onto syntactic entities in a principled manner. For instance, the linking principle in conceptual semantics states that the least embedded argument in the CS is mapped onto the least embedded syntactic argument, namely the subject. In (22b & c), [JOHN] is the least embedded argument in both the action and thematic tiers; this explains why [JOHN] instead of [BALL] is mapped onto the subject of (22a). Jun (2003) is a conceptual semantic work on case; Culicover and Jackendoff (2005) offer conceptual semantic treatments of binding, control, and many other syntax-related phenomena. In short, conceptual semantics is an interface theory between syntax and semantics. The theory has a desirable consequence for the learnability problem, too. Language learners cannot acquire language solely by syntax or solely by semantics. As Levin (1993) demonstrates, a number of syntactic regularities are predicted by semantic properties of predicates. Conceptual semantics makes a number of predictions about syntax in terms of CS. Chomsky's explanatory adequacy is a requirement for the learnability problem; conceptual semantics is thus a theory that aims to achieve the highest goal of a linguistic theory.

The final requirement on a semantic theory is the cognitive constraint: a semantic theory should address interface problems between language and other peripheral systems like vision, hearing, smell, taste, kinesthesia, etc. Conceptual semantics fulfills this requirement, as CS is by definition a level of mental representation at which both linguistic and non-linguistic modalities converge. Jackendoff (1987c) focuses on the interface problem, and shows, for instance, how the visual representation is formally compatible with the linguistic representation based on Marr's (1982) theory of visual perception.

Comparison with Other Works

Bierwisch and Schreuder's (B&S; 1992) work is another influential theory that makes explicit use of

the term *conceptual structure*. Conceptual semantics shares two important assumptions with B&S, but there are crucial distinctions between the two theories. First, B&S also assume a separate level of conceptual structure. Their conception of CS is similar to Jackendoff's conception of CS in that CS is a representational system of message structure where non-linguistic factual/encyclopedic information is expressed. B&S, however, assume that CS strictly belongs to a nonlinguistic modality, and that the linguistic meaning is represented in another level called *semantic form* (SF). As a result, SF, but not CS, is the object of lexical semantics, and hence LCS does not make much sense in this theory. In the first section of this article, we discussed two possible views of CS; B&S take the former view of CS, whereas Jackendoff advocates the latter view.

Second, SF in B&S's theory is compositional as CS in conceptual semantics. B&S's lexical decomposition relies on two sorts of elements: constants such as DO, MOVE, FIN, LOC, etc., and variables such as *x*, *y*, *z*. Constants and variables are composed into a larger expression in terms of formal logic. (23a) illustrates B&S's SF for *enter*; (23b) is the CS for the same word in Jackendoff's theory.

(23a) [*y* DO [MOVE *y*] : FIN [*y* LOC IN *x*]]

(23b) [_{Event} GO ([_{Thing}], [_{Path} TO ([_{Place} IN ([_{Thing}]))])])]

One reason B&S maintain a purely nonlinguistic CS as well as a separate SF is that factual or encyclopedic knowledge does not seem to make much grammatical contribution to language. To B&S, there is a clear boundary where the *semantic* and the *encyclopedic* diverge.

Pustejovsky's (1995) generative lexicon (GL) theory is interesting in this regard. GL also assumes lexical decomposition. Pustejovsky's lexical decomposition makes use of factual or encyclopedic knowledge in a rigorous formalism called the *qualia* structure. The qualia structure of *book*, for instance, expresses such factual knowledge as the origin of *book* as *write*(*x*, *y*) in the Agentive quale, where *x* is a writer (i.e., *human*(*x*)), and *y* is a book (i.e., *book*(*y*)). The qualia structure also expresses the use of the word in the Telic quale; hence, the lexical semantic structure for *book* includes such factual knowledge as *read*(*w*, *y*), where *w* is a reader (i.e., *human*(*w*)), and *y* is a book.

The factual or encyclopedic knowledge is not only expressed in formal linguistic representations but also plays a crucial role in explaining a significant portion of linguistic phenomena. We interpret (24) as either *Chomsky began writing a book* or *Chomsky began reading a book*. Pustejovsky suggests generative

devices like type coercion and co-composition to explain the two readings of (24) in a formal theory; i.e., *writing* or *reading* is part of the qualia structure of *book*, and, hence, the two readings of (24) are predicted by formal principles of lexical semantics.

(24) Chomsky began a book.

It is far beyond the scope of this article to discuss the GL theory in detail. But the success of the GL theory for a vast range of empirical data shows that the boundary between *semantic* and *encyclopedic* or between *linguistic* and *nonlinguistic* is not so clear as B&S assume in their distinction between CS and SF.

Suggested Readings

For a quick overview of conceptual semantics with one paper, see Jackendoff (1987a). For foundational issues of conceptual semantics, see Jackendoff (1983). For an overview of language and other cognitive capacities from a broad perspective, see Jackendoff (1987c). Jackendoff (1990) offers a comprehensive picture of conceptual semantics. Jackendoff (1997) is a bit technical, but it is important to set up the parallel architecture of language. For syntactic issues of conceptual semantics, see Jun (2003) and Culicover and Jackendoff (2005).

Bibliography

- Berko Gleason J (ed.) (1997). *The development of language*. Boston: Allyn and Bacon.
- Bierwisch M & Schreuder R (1992). 'From concepts to lexical items.' *Cognition* 42, 23–60.
- Chomsky N (1970). 'Remarks on nominalization.' In Jacobs R A & Rosenbaum P S (eds.) *Readings in English Transformational Grammar*. Waltham: Ginn and Company. 184–221.
- Chomsky N (1981). *Lectures on government and binding: the Pisa lectures*. Dordrecht: Foris.
- Chomsky N (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Csuri P (1996). 'Generalized dependencies: description, reference, and anaphora.' Ph.D. diss., Brandeis University.
- Collins A & Quillian M (1969). 'Retrieval time from semantic memory.' *Journal of Verbal Learning and Verbal Behavior* 9, 240–247.
- Culicover P & Jackendoff R (2005). *Simpler syntax*. Oxford: Oxford Univ. Press.
- Fodor J A (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Jackendoff R (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Jackendoff R (1977). *X-bar syntax: a study of phrase structure*. Cambridge, MA: MIT Press.
- Jackendoff R (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff R (1987a). 'The status of thematic relations in linguistic theory.' *Linguistic Inquiry* 18, 369–411.
- Jackendoff R (1987b). 'X-bar semantics.' In Pustejovsky James (ed.) *Semantics and the lexicon*. Dordrecht: Kluwer Academic Publishers. 15–26.
- Jackendoff R (1987c). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.
- Jackendoff R (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Jackendoff R (1991). 'Parts and boundaries.' *Cognition* 41, 9–45.
- Jackendoff R (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff R (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Jun J S (2003). 'Syntactic and semantic bases of case assignment: a study of verbal nouns, light verbs, and dative.' Ph.D. diss., Brandeis University.
- Katz J J (1980). 'Chomsky on meaning.' *Language* 56(1), 1–41.
- Katz J J & Fodor J A (1963). 'The structure of a semantic theory.' *Language* 39(2), 170–210.
- Labov W (1973). 'The boundaries of words and their meanings.' In Bailey C -J N & Shuy R W (eds.) *New ways of analyzing variation in English*, vol. 1. Washington, DC: Georgetown University Press.
- Levin B (1993). *English verb classes and alternations*. Chicago: University of Chicago Press.
- Marr D (1982). *Vision*. San Francisco: W. H. Freeman.
- Piñango M M, Zurif E & Jackendoff R (1999). 'Real-time processing implications of aspectual coercion at the syntax-semantics interface.' *Journal of Psycholinguistic Research* 28(4), 395–414.
- Pustejovsky J (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Swinney D (1979). 'Lexical access during sentence comprehension: (re)consideration of context effects.' *Journal of Verbal Learning and Verbal Behavior* 18, 645–659.
- Wertheimer M (1912). 'Experimentelle Studien über das Sehen von Bewegung.' *Zeitschrift für Psychologie* 61, 161–265.
- Zurif E & Blumstein S (1978). 'Language and the brain.' In Halle M, Bresnan J & Miller G A (eds.) *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press. 229–245.
- Zurif E, Caramazza A & Myerson R (1972). 'Grammatical judgments of agrammatic aphasics.' *Neuropsychologia* 10, 405–417.

Lexical Semantics: Overview

J Pustejovsky, Brandeis University, Waltham, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Word Knowledge

Semantic interpretation requires access to knowledge about words. The lexicon of a grammar must provide a systematic and efficient way of encoding the information associated with words in a language. Lexical semantics is the study of what words mean and how they structure these meanings. This article examines word meaning from two different perspectives: the information required for composition in the syntax and the knowledge needed for semantic interpretation. The lexicon is not merely a collection of words with their associated phonetic, orthographic, and semantic forms. Rather, lexical entries are structured objects that participate in larger operations and compositions, both enabling syntactic environments and acting as signatures to semantic entailments and implicatures in the context of larger discourse.

There are four basic questions in modeling the semantic content and structure of the lexicon: (1) What semantic information goes into a lexical entry? (2) How do lexical entries relate semantically to one another? (3) How is this information exploited compositionally by the grammar? and (4) How is this information available to semantic interpretation generally? This article focuses on the first two.

The lexicon and lexical semantics have traditionally been viewed as the most passive modules of language, acting in the service of the more dynamic components of the grammar. This view has its origins in the generative tradition (Chomsky, 1955) and has been an integral part of the notion of the lexicon ever since. While the *aspects* model of selectional features (Chomsky, 1965) restricted the relation of selection to that between lexical items, work by McCawley (1968) and Jackendoff (1972) showed that selectional restrictions must be available to computations at the level of derived semantic representation rather than at deep structure. Subsequent work by Bresnan (1982), Gazdar *et al.* (1985), and Pollard and Sag (1994) extended the range of phenomena that can be handled by the projection and exploitation of lexically derived information in the grammar.

With the convergence of several areas in linguistics (lexical semantics, computational lexicons, and type theories) several models for the determination of selection have emerged that put even more compositional power in the lexicon, making explicit reference to the paradigmatic systems that allow for grammatical

constructions to be partially determined by selection. Examples of this approach are generative lexicon theory (Bouillon and Busa, 2001; Pustejovsky, 1995) and construction grammar (Goldberg, 1995; Jackendoff, 1997, 2002). These developments have helped to characterize the approaches to lexical design in terms of a hierarchy of semantic expressiveness. There are at least three such classes of lexical description: *sense enumerative lexicons*, where lexical items have a single type and meaning, and ambiguity is treated by multiple listings of words; *polymorphic lexicons*, where lexical items are active objects, contributing to the determination of meaning in context, under well-defined constraints; and *unrestricted sense lexicons*, where the meanings of lexical items are determined mostly by context and conventional use. Clearly, the most promising direction seems to be a careful and formal elucidation of the *polymorphic lexicons*, and this will form the basis of the subsequent discussion of both the structure and the content of lexical entries.

Historical Overview

The study of word meaning has occupied philosophers for centuries, beginning at least with Aristotle's theory of meaning. Locke, Hume, and Reid all paid particular attention to the meanings of words, but not until the 19th century did the rise of philological and psychological investigations of word meaning occur, with Bréal (1897), Erdmann (1900), Trier (1931), Stern (1931/1968), and others focused on word connotation, semantic drift, and word associations in the mental lexicon as well as in social contexts.

Interestingly, Russell, Frege, and other early analytic philosophers were not interested in language as a linguistic phenomenon but simply as the medium through which judgments can be formed and expressed. Hence, there is little regard for the relations between senses of words, when not affecting the nature of judgment, for example, within intensional contexts. Nineteenth-century semanticists and semasiologists, on the other hand, viewed polysemy as the life force of human language. Bréal, for example, considered it to be a necessary creative component of language and argued that this phenomenon better than most in semantics illustrates the cognitive and conceptualizing force of the human species.

Even with their obvious enthusiasm, semasiology produced no lasting legacy to the study of lexical semantics. In fact, there was no systematic research into lexical meaning until structural linguists extended the relational techniques of Saussure (1916/1983) and elaborated the framework of componential analysis for language meaning (Jakobson, 1970).

The idea behind componential analysis is the reduction of a word's meaning into its ultimate contrastive elements. These contrastive elements are structured in a matrix, allowing for dimensional analysis and generalizations to be made about lexical sets occupying the cells in the matrix.

This technique developed into a general framework for linguistic description called distinctive feature analysis (Jakobson and Halle, 1956). This is essentially the inspiration for Katz and Fodor's 1963 theory of lexical semantics within transformational grammar. In this theory, usually referred to as 'markerese,' a lexical entry in the language consists of grammatical and semantic markers and a special feature called a '*semantic distinguisher*.' In Weinreich (1972) and much subsequent discussion, it was demonstrated that this model is far too impoverished to characterize the compositional mechanisms inherent in language. In the late 1960s and early 1970s, alternative models of word meaning emerged (Fillmore, 1965; Gruber, 1965; Jackendoff, 1972; Lakoff, 1965) that respected the relational structure of sentence meaning while encoding the named semantic functions in lexical entries. In Dowty (1979), a model theoretic interpretation of the decompositional techniques of Lakoff, McCawley, and Ross was developed.

Recently, the role of lexical-syntactic mapping has become more evident, particularly with the growing concern over projection from lexical semantic form, the problem of verbal alternations and polyvalency, and the phenomenon of polysemy.

Ambiguity and Polysemy

Given the compactness of a lexicon relative to the number of objects and relations in the world, and the concepts we have for them, lexical ambiguity is inevitable. Add to this the cultural, historical, and linguistic blending that contributes to the meanings of our lexical items, and ambiguity can appear arbitrary as well. Hence, '*homonymy*' – where one lexical form has many meanings – is to be expected in a language. Examples of homonyms are illustrated in the following sentences:

- (1a) Mary walked along the bank of the river.
- (1b) Bank of America is the largest bank in the city.
- (2a) Drop me a line when you are in Boston.
- (2b) We built a fence along the property line.
- (3a) First we leave the gate, then we taxi down the runway.
- (3b) John saw the taxi on the street.
- (4a) The judge asked the defendant to approach the bar.
- (4b) The defendant was in the pub at the bar.

Weinreich (1964) calls such lexical distinctions '*contrastive ambiguity*,' where it is clear that the senses associated with the lexical item are unrelated. For this reason, it is generally assumed that homonyms are represented as separate lexical entries within the organization of the lexicon. This accords with a view of lexical organization that has been termed a '*sense enumeration lexicon*' (cf. Pustejovsky, 1995). That is, a lexicon is *sense enumerative* when every word *w* that has multiple senses stores these senses as separate lexical entries.

This model becomes difficult to maintain, however, when we consider the phenomenon known as '*polysemy*.' Polysemy is the relationship that exists between different senses of a word that are related in some logical manner rather than arbitrarily, as in the previous examples. We can distinguish three broad types of polysemy, each presenting a novel set of challenges to lexical semantics and linguistic theory.

- a. Deep semantic typing: single argument polymorphism
- b. Syntactic alternations: multiple argument polymorphism
- c. Dot objects: lexical reference to objects that have multiple facets

The first class refers mainly to functors allowing a range of syntactic variation in a single argument. For example, aspectual verbs (*begin* and *finish*), perception verbs (*see*, *hear*), and most propositional attitude verbs (*know*, *believe*) subcategorize for multiple syntactic forms in complement position, as illustrated in (6):

- (5a) Mary began to read the novel.
- (5b) Mary began reading the novel.
- (5c) Mary began the novel.
- (6a) Bill saw John leave.
- (6b) Bill saw John leaving.
- (6c) Bill saw John.
- (7a) Mary believes that John told the truth.
- (7b) Mary believes what John said.
- (7c) Mary believes John's story.

What these and many other cases of multiple selection share is that the underlying relation between the verb and each of its complements is essentially identical. For example, in (7), the complement to the verb *believe* in all three sentences is a proposition; in (5), what is begun in each sentence is an event of some sort; and in (6), the object of the perception is (arguably) an event in each case. This has led some linguists to argue for semantic selection (cf. Chomsky, 1986; Grimshaw, 1979) and others to argue for structured

selectional inheritance (Godard and Jayez, 1993). In fact, these perspectives are not that distant from one another (cf. Pustejovsky, 1995): in either view, there is an explicit lexical association between syntactic forms that is formally modeled by the grammar.

The second type of polysemy (syntactic alternations) involves verbal forms taking arguments in alternating constructions, the so-called '*verbal alternations*' (cf. Levin, 1993). These are true instances of polysemy because there is a logical (typically *causal*) relation between the two senses of the verb. As a result, the lexicon must either relate the senses through lexical rules (such as in head-driven phrase structure grammar (HPSG) treatments; cf. Pollard and Sag, 1994) or assume that there is one lexical form that has multiple syntactic realizations (cf. Pustejovsky and Busa, 1995).

(8a) The window opened suddenly.

(8b) Mary opened the window suddenly.

(9a) Bill began his lecture on time.

(9b) The lecture began on time.

(10a) The milk spilled onto the table.

(10b) Mary spilled the milk onto the table.

The final form of polysemy reviewed here is encountered mostly in nominals and has been termed '*regular polysemy*' (cf. Apresjan, 1973) and '*logical polysemy*' (cf. Pustejovsky, 1991) in the literature; it is illustrated in the following sentences:

(11a) Mary carried the book home.

(11b) Mary doesn't agree with the book.

(12a) Mary has her lunch in her backpack.

(12b) Lunch was longer today than it was yesterday.

(13a) The flight lasted 3 hours.

(13b) The flight landed on time in Los Angeles.

Notice that in each of the pairs, the same nominal form is assuming different semantic interpretations relative to its selective context. For example, in (11a) the noun *book* refers to a physical object, whereas in (11b) it refers to the informational content. In (12a), *lunch* refers to the physical manifestation of the food, whereas in (12b) it refers to the eating event. Finally, in (13a) *flight* refers to the flying event, whereas in (13b) it refers to the plane. This phenomenon of polysemy is one of the most challenging in the area and has stimulated much research (Bouillon, 1997; Bouillon and Busa, 2001). In order to understand how each of these cases of polysemy can be handled, we must first familiarize ourselves with the structure of individual lexical entries.

Lexical Relations

Another important aspect of lexical semantics is the study of how words are semantically related to one another. Four classes of lexical relations, in particular, are important to recognize: synonymy, antonymy, hyponymy, and meronymy.

Synonymy is generally taken to be a relation between words rather than concepts. One fairly standard definition states that two expressions are synonymous if substituting one for the other in all contexts does not change the truth value of the sentence where the substitution is made (cf. Cruse, 1986, 2004; Lyons, 1977). A somewhat weaker definition makes reference to the substitution relative to a specific context. For example, in the context of carpentry, *plank* and *board* might be considered synonyms, but not necessarily in other domains (cf. Miller *et al.*, 1990). The relation of antonymy is characterized in terms of semantic opposition and, like synonymy, is properly defined over pairs of lexical items rather than concepts. Examples of antonymy are *rise/fall*, *heavy/light*, *fast/slow*, and *long/short* (cf. Cruse, 1986; Miller, 1991). It is interesting to observe that co-occurrence data illustrate that synonyms do not necessarily share the same antonyms. For example, *rise* and *ascend* as well as *fall* and *descend* are similar in meaning, yet neither *fall/ascend* nor *rise/descend* are antonym pairs. For further details see Miller *et al.* (1990).

The most studied relation in the lexical semantic community is hyponymy, the taxonomic relationship between words, as defined in WordNet (Fellbaum, 1998) and other semantic networks. For example, specifying *car* as a hyponym of *vehicle* is equivalent to saying that *vehicle* is a superconcept of the concept *car* or that the set *car* is a subset of those individuals denoted by the set *vehicle*.

One of the most difficult lexical relations to define and treat formally is that of meronymy, the relation of parts to the whole. The relation is familiar from knowledge representation languages with predicates or slot-names such as 'part-of' and 'made-of.' For treatments of this relation in lexical semantics, see Miller *et al.* (1990) and Cruse (1986).

The Semantics of a Lexical Entry

It is generally assumed that there are four components to a lexical item: phonological, orthographic, syntactic, and semantic information. Here, we focus first on syntactic features and then on what semantic information must be encoded in an individual lexical entry.

There are two types of syntactic knowledge associated with a lexical item: its *category* and its *subcategory*. The former includes traditional classifications of

both the major categories, such as noun, verb, adjective, adverb, and preposition, as well as the minor categories, such as adverbs, conjunctions, quantifier elements, and determiners. Knowledge of the subcategory of a lexical item is typically information that differentiates categories into distinct, distributional classes. This sort of information may be usefully separated into two types, *contextual features* and *inherent features*. The former are features that may be defined in terms of the contexts in which a given lexical entry may occur. Subcategorization information marks the local syntactic context for a word. It is this information that ensures that the verb *devour*, for example, is always transitive in English, requiring a direct object; the lexical entry encodes this requirement with a subcategorization feature specifying that a noun phrase (NP) appear to its right. Another type of context encoding is collocational information, where patterns that are not fully productive in the grammar can be tagged. For example, the adjective *heavy* as applied to *drinker* and *smoker* is collocational and not freely productive in the language (Mel'čuk, 1988). '*Inherent features*' on the other hand, are properties of lexical entries that are not easily reduced to a contextual definition but, rather, refer to the ontological typing of an entity. These include such features as count/mass (e.g., *pebble* vs. *water*), abstract, animate, human, physical, and so on.

Lexical items can be systematically grouped according to their syntactic and semantic behavior in the language. For this reason, there have been two major traditions of word clustering, corresponding to this distinction. Broadly speaking, for those concerned mainly with grammatical behavior, the most salient aspect of a lexical item is its *argument structure*; for those focusing on a word's entailment properties, the most important aspect is its *semantic class*. In this section, these two approaches are examined and it is shown how their concerns can be integrated into a common lexical representation.

Lexical Semantic Classifications

Conventional approaches to lexicon design and lexicography have been relatively informal with regard to forming taxonomic structures for the word senses in the language. For example, the top concepts in WordNet (Miller *et al.*, 1990) illustrate how words are characterized by local clusterings of semantic properties. As with many ontologies, however, it is difficult to discern a coherent global structure for the resulting classification beyond a weak descriptive labeling of words into extensionally defined sets.

One of the most common ways to organize lexical knowledge is by means of type or feature inheritance mechanisms (Carpenter, 1992; Copestake and

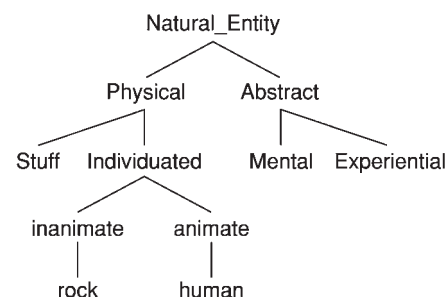


Figure 1 Type structures.

Briscoe, 1992; Evans and Gazdar, 1990; Pollard and Sag, 1994). Furthermore, Briscoe *et al.* (1993) described a rich system of types for allowing default mechanisms into lexical type descriptions. Similarly, type structures, such as that shown in **Figure 1**, can express the inheritance of syntactic and semantic features, as well as the relationship between syntactic classes and alternations (cf. Alsina, 1992; Davis, 1996; Koenig and Davis, 1999; Sanfilippo, 1993) and other relations (cf. Pustejovsky, 2001; Pustejovsky and Boguraev, 1993).

In the remainder of this section, we first examine the approach to characterizing the weak constraints imposed on a lexical item associated with its arguments. Then, we examine attempts to model lexical behavior by means of internal constraints imposed on the predicate. Finally, it is shown how, in some respects, these are very similar enterprises and both sets of constraints may be necessary to model lexical behavior.

Argument Structure

Once the base syntactic and semantic typing for a lexical item has been specified, its subcategorization and selectional information must be encoded in some form. There are two major techniques for representing this type of knowledge:

1. Associate 'named roles' with the arguments to the lexical item (Fillmore, 1985; Gruber, 1965; Jackendoff, 1972).
2. Associate a logical decomposition with the lexical item; meanings of arguments are determined by how the structural properties of the representation are interpreted (cf. Hale and Keyser, 1993; Jackendoff, 1983; Levin and Rappaport, 1995).

One influential way of encoding selectional behavior has been the theory of *thematic relations* (cf. Gruber, 1976; Jackendoff, 1972). Thematic relations are now generally defined as partial semantic functions of the event being denoted by the verb or noun, and they behave according to a predefined calculus of roles relations (e.g., Dowty, 1989). For example,

semantic roles such as agent, theme, and goal can be used to partially determine the meaning of a predicate when they are associated with the grammatical arguments to a verb.

- (14a) **put** <AGENT, THEME, LOCATION>
 (14b) **borrow** <RECIPIENT, THEME, SOURCE>

Thematic roles can be ordered relative to each other in terms of an implicational hierarchy. For example, there is considerable use of a universal subject hierarchy such as is shown in the following (cf. Comrie, 1981; Fillmore, 1968):

- (15) AGENT > RECIPIENT/BENEFACTIVE >
 THEME/PATIENT > INSTRUMENT >
 LOCATION>

Many linguists have questioned the general explanatory coverage of thematic roles, however, and have chosen alternative methods for capturing the generalizations they promised. Dowty (1991) suggested that theta-role generalizations are best captured by entailments associated with the predicate. A theta-role can then be seen as the set of predicate entailments that are properties of a particular argument to the verb. Characteristic entailments might be thought of as prototype roles, or proto-roles; this allows for degrees or shades of meaning associated with the arguments to a predicate. Others have opted for a more semantically neutral set of labels to assign to the parameters of a relation, whether it is realized as a verb, noun, or adjective. For example, the theory of argument structure as developed by Williams (1981), Grimshaw (1990), and others can be seen as a move toward a more minimalist description of semantic differentiation in the verb's list of parameters. The argument structure for a word can be seen as the simplest specification of its semantics, indicating the number and type of parameters associated with the lexical item as a predicate. For example, the verb *die* can be represented as a predicate taking one argument, *kill* as taking two arguments, where as the verb *give* takes three arguments:

- (16a) **die** (x)
 (16b) **kill** (x,y)
 (16c) **give** (x,y,z)

What originally began as the simple listing of the parameters or arguments associated with a predicate has developed into a sophisticated view of the way arguments are mapped onto syntactic expressions. Williams's (1981) distinction between *external* (the underlined arguments above) and *internal* arguments and Grimshaw's proposal for a hierarchically structured representation (cf. Grimshaw, 1990) provide us with the basic syntax for one aspect of a word's

meaning. Similar remarks hold for the argument list structure in HPSG (Pollard and Sag, 1994) and Lexical Functional Grammar (Bresnan, 1994).

The interaction of a structured argument list and a rich system of types, such as that presented previously, provides a mechanism for semantic selection through inheritance. Consider, for instance, the sentence pairs in (17):

- (17a) The man/the rock fell.
 (17b) The man/*the rock died.

Now consider how the selectional distinction for a feature such as animacy is modeled so as to explain the selectional constraints of predicates. For the purpose of illustration, the arguments of a verb will be identified as being typed from the system shown previously.

- (18a) λx : physical[**fall**(x)]
 (18b) λx : animate[**die**(x)]

In the sentences in (17), it is clear how rocks cannot die and men can, but it is still not obvious how this judgment is computed, given what we would assume are the types associated with the nouns *rock* and *man*, respectively. What accomplishes this computation is a rule of subtyping, Θ , that allows the type associated with the noun *man* (i.e., 'human') to also be accepted as the type 'animate,' which is what the predicate *die* requires of its argument as stated in (18b) (cf. Carpenter, 1992).

- (19) Θ [*human* \sqsubseteq *animate*]: *human* \rightarrow *animate*

The rule Θ , applies since the concept 'human' is subtyped under 'animate' in the type hierarchy. Parallel considerations rule out the noun *rock* as a legitimate argument to *die* since it is not subtyped under 'animate.' Hence, one of the concerns given previously for how syntactic processes can systematically keep track of which 'selectional features' are entailed and which are not is partially addressed by such lattice traversal rules as the one presented here.

Event Structure and Lexical Decomposition

The second approach to lexical specification mentioned previously is to define constraints internally to the predicate. Traditionally, this has been known as 'lexical decomposition.' In this section, we review the motivations for decomposition in linguistic theory and the proposals for encoding lexical knowledge as structured objects. We then relate this to the way in which verbs can be decomposed in terms of eventualities (Tenny and Pustejovsky, 2000).

Since the 1960s, lexical semanticists have attempted to formally model the semantic relations between

lexical items such as between the adjective *dead* and the verbs *die* and *kill* (cf. Lakoff, 1965; McCawley, 1968) in the following sentences:

- (20a) John killed Bill.
- (20b) Bill died.
- (20c) Bill is dead.

Assuming the underlying form for a verb such as *kill* directly encodes the stative predicate in (20c) and the relation of causation, generative semanticists posited representations such as (21):

- (21) (CAUSE (x , (BECOME (NOT (ALIVE y))))

Here the predicate CAUSE is represented as a relation between an individual causer x and an expression involving a change of state in the argument y . Carter (1976) proposes a representation quite similar, shown here for the causative verb *darken*:

- (22) (x CAUSE ((y BE DARK) CHANGE))

Although there is an intuition that the cause relation involves a causer and an event, neither Lakoff nor Carter make this commitment explicitly. In fact, it has taken several decades for Davidson's (1967) observations regarding the role of events in the determination of verb meaning to find their way convincingly into the major linguistic frameworks. A new synthesis has emerged that attempts to model verb meanings as complex predicative structures with rich event structures (cf. Hale and Keyser, 1993; Parsons, 1990; Pustejovsky, 1991). This research has developed the idea that the meaning of a verb can be analyzed into a structured representation of the event that the verb designates, and it has furthermore contributed to the realization that verbs may have complex, internal event structures. Recent work has converged on the view that complex events are structured into an inner and an outer event, where the outer event is associated with causation and agency, and the inner event is associated with telicity (completion) and change of state (cf. Tenny and Pustejovsky, 2000).

Jackendoff (1990) developed an extensive system of what he calls '*conceptual representations*,' which parallel the syntactic representations of sentences of natural language. These employ a set of canonical predicates, including CAUSE, GO, TO, and ON, and canonical elements, including Thing, Path, and Event. These approaches represent verb meaning by decomposing the predicate into more basic predicates. This work owes obvious debt to the innovative work within generative semantics, as illustrated by McCawley's (1968) analysis of the verb *kill*. Recent versions of lexical representations inspired by

generative semantics can be seen in the lexical relational structures of Hale and Keyser (1993), where syntactic tree structures are employed to capture the same elements of causation and change of state as in the representations of Carter, Levin and Rapoport, Jackendoff, and Dowty. The work of Levin and Rappaport, building on Jackendoff's lexical conceptual structures, has been influential in further articulating the internal structure of verb meanings (Levin and Rappaport, 1995).

Pustejovsky (1991) extended the decompositional approach presented in Dowty (1979) by explicitly reifying the events and subevents in the predicative expressions. Unlike Dowty's treatment of lexical semantics, where the decompositional calculus builds on propositional or predicative units (as discussed previously) a 'syntax of event structure' makes explicit reference to quantified events as part of the word meaning. Pustejovsky further introduced a tree structure to represent the temporal ordering and dominance constraints on an event and its subevents. For example, a predicate such as *build* is associated with a complex event such as the following (cf. also Moens and Steedman, 1988):

- (23) [*transition*[e_1 :PROCESS] [e_2 :STATE]]

The process consists of the building activity itself, whereas the state represents the result of there being the object built. Grimshaw (1990) adopted this theory in her work on argument structure, where complex events such as *break* are given a similar representation. In such structures, the process consists of what x does to cause the breaking, and the state is the resultant state of the broken item. The process corresponds to the outer causing event as discussed previously, and the state corresponds in part to the inner change of state event. Both Pustejovsky and Grimshaw differ from the previous authors in assuming a specific level of representation for event structure, distinct from the representation of other lexical properties. Furthermore, they follow Higginbotham (1985) in adopting an explicit reference to the event place in the verbal semantics.

Rappaport and Levin (2001) adopted a large component of the event structure model for their analysis of the resultative construction in English. Event decomposition has also been employed for properties of adjectival selection, the interpretation of compounds, and stage and individual-level predication.

Qualia Structure

Thus far, we have focused on the lexical semantics of verb entries. All of the major categories, however,

are encoded with syntactic and semantic feature structures that determine their constructional behavior and subsequent meaning at logical form. In generative lexicon theory (Pustejovsky, 1995), it is assumed that word meaning is structured on the basis of four generative factors, or ‘*qualia roles*’, that capture how humans understand objects and relations in the world and provide the minimal explanation for the linguistic behavior of lexical items (these are inspired in large part by Moravcsik’s (1975, 1990) interpretation of Aristotelian *aitia*). These are the formal role, the basic category that distinguishes the object within a larger domain; the constitutive role, the relation between an object and its constituent parts; the telic role, its purpose and function; and the agentive role, factors involved in the object’s origin or ‘coming into being.’ Qualia structure is at the core of the generative properties of the lexicon since it provides a general strategy for creating new types. For example, consider the properties of nouns such as *rock* and *chair*. These nouns can be distinguished on the basis of semantic criteria that classify them in terms of general categories such as *natural_kind*, *artifact_object*. Although very useful, this is not sufficient to discriminate semantic types in a way that also accounts for their grammatical behavior. A crucial distinction between *rock* and *chair* concerns the properties that differentiate *natural_kinds* from *artifacts*: Functionality plays a crucial role in the process of individuation of artifacts but not of natural kinds. This is reflected in grammatical behavior, whereby ‘a good chair’ or ‘enjoy the chair’ are well-formed expressions reflecting the specific purpose for which an artifact is designed, but ‘good rock’ or ‘enjoy a rock’ are semantically ill formed since for *rock* the functionality (i.e., telic) is undefined. Exceptions exist when new concepts are referred to, such as when the object is construed relative to a specific activity, as in ‘The climber enjoyed that rock’; *rock* takes on a new meaning by virtue of having telicity associated with it, and this is accomplished by integration with the semantics of the subject NP. Although *chair* and *rock* are both *physical_object*, they differ in their mode of coming into being (i.e., agentive): artifacts are man-made, *rocks* develop in nature. Similarly, a concept such as *food* or *cookie* has a physical manifestation or denotation, but also a functional grounding, pertaining to the relation of ‘eating.’ These apparently contradictory aspects of a category are orthogonally represented by the qualia structure for that concept, which provides a coherent structuring for different dimensions of meaning.

See also: Compositionality: Semantic Aspects; Syntax-Semantics Interface.

Bibliography

- Alsina A (1992). ‘On the argument structure of causatives.’ *Linguistic Inquiry* 23(4), 517–555.
- Apresjan J D (1973). ‘Regular polysemy.’ *Linguistics* 142, 5–32.
- Bouillon P (1998). *Polymorphie et sémantique lexicale: le cas des adjectifs*. Lille: Presse du Septentrion.
- Bouillon P & Busa F (2001). *The language of word meaning*. Cambridge [England], New York: Cambridge University Press.
- Bréal M (1897). *Essai de sémantique (science des significations)*. Paris: Hachette.
- Bresnan J (ed.) (1982). *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.
- Bresnan J (1994). ‘Locative inversion and the architecture of universal grammar.’ *Language* 70, 72–131.
- Briscoe T, de Paiva V & Copestake A (eds.) (1993). *Inheritance, defaults, and the lexicon*. Cambridge, UK: Cambridge University Press.
- Carpenter R (1992). ‘Typed feature structures.’ *Computational Linguistics* 18, 2.
- Chomsky N (1955). *The logical structure of linguistic theory*. Chicago: University of Chicago Press (Original work published 1975).
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Comrie B (1981). *Language universals and linguistic typology*. Chicago, IL: The University of Chicago Press.
- Copestake A & Briscoe T (1992). ‘Lexical operations in a unification-based framework.’ In Pustejovsky J & Bergler S (eds.) *Lexical semantics and knowledge representation*, Berlin: Springer Verlag.
- Cruse D A (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.
- Cruse D A (2004). *Meaning in language: an introduction to semantics and pragmatics* (2nd edn.). Oxford: Oxford University Press.
- Davidson D (1967). ‘The logical form of action sentences.’ In Rescher N (ed.) *The logic of decision and action*. Pittsburgh: Pittsburgh University Press.
- Davis A (1996). *Lexical semantics and linking and the hierarchical lexicon*. Ph.D. diss., Stanford University.
- Davis A & Koenig J-P (2000). ‘Linking as constraints on word classes in a hierarchical lexicon.’ *Language* 2000.
- Dowty D R (1979). *Word meaning and Montague grammar*. Dordrecht, The Netherlands: D. Reidel.
- Dowty D R (1989). ‘On the semantic content of the notion “thematic role”.’ In Chierchia G, Partee B & Turner R (eds.) *Properties, types, and meaning*, vol. 2. Semantic issues. Dordrecht: D. Reidel.
- Dowty D (1991). ‘Thematic proto-roles and argument selection.’ *Language* 67, 547–619.
- Erdmann K (1900). *Die Bedeutung des Wortes: Aufsätze aus dem Grenzgebiet der Sprachpsychologie und Logik*. Avenarius: Leipzig.

- Evans R & Gazdar G (1990). 'The DATR papers: February 1990.' *Cognitive Science Research Paper CSRP 139*, School of Cognitive and Computing Science, University of Sussex, Brighton, England.
- Fillmore C (1965). *Entailment rules in a semantic theory*. POLA Report 10. Columbus, OH: Ohio State University.
- Fillmore C (1968). 'The case for case.' In Bach E W & Harms R T (eds.) *Universals in linguistic theory*. New York: Holt, Rinehart and Winston.
- Gazdar G, Klein E, Pullum G & Sag I (1985). *Generalized phrase structure grammar*. Cambridge, MA: Harvard University Press.
- Goldberg A (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Grimshaw J (1979). 'Complement selection and the lexicon.' *Linguistic Inquiry* 10, 279–326.
- Grimshaw J (1990). *Argument structure*. Cambridge: MIT Press.
- Gruber J S (1965/1976). *Lexical structures in syntax and semantics*. Amsterdam: North-Holland.
- Gruber J S (1976). *Lexical structures in syntax and semantics*. Amsterdam: North-Holland.
- Hale K & Keyser J (1993). *On argument structure and the lexical expression of syntactic relations: the view from building 20*. Cambridge, MA: MIT Press.
- Halle M, Bresnan J & Miller G (eds.) (1978). *Linguistic theory and psychological reality*. Cambridge: MIT Press.
- Higginbotham J (1985). 'On Semantics.' *Linguistic Inquiry* 16, 547–593.
- Hjelmslev L (1961). *Prolegomena to a theory of language*. Whitfield F (ed.). Madison: University of Wisconsin Press (Original work published 1943).
- Jackendoff R (1972). *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- Jackendoff R (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff R (1990). *Semantic structures*. Cambridge: MIT Press.
- Jackendoff R (1992). 'Babe Ruth homered his way into the hearts of America.' In Stowell T & Wehrli E (eds.) *Syntax and the lexicon*. San Diego: Academic Press. 155–178.
- Jackendoff R (2002). *Foundations of language: brain, meaning, grammar*. Oxford: Oxford University Press.
- Jakobson R (1970). *Recent developments in linguistic science*. Perennial Press.
- Jakobson R (1974). *Main trends in the science of language*. New York: Harper & Row.
- Jakobson R & Halle M (1956). *Fundamentals of language*. The Hague, The Netherlands: Mouton.
- Katz J (1972). *Semantic theory*. New York: Harper & Row.
- Katz J & Fodor J (1963). 'The structure of a semantic theory.' *Language* 39, 170–210.
- Lakoff G (1965/1970). *Irregularity in syntax*. New York: Holt, Rinehart, and Winston.
- Levin B & Rappaport Hovav M (1995). *Unaccusativity: at the syntax-semantics interface*. Cambridge: MIT Press.
- Lyons J (1977). *Semantics* (2 volumes). Cambridge: Cambridge University Press.
- McCawley J (1968). 'Lexical insertion in a transformational grammar without deep structure.' *Proceedings of the Chicago Linguistic Society* 4.
- Mel'cuk I A (1988b). *Dependency syntax*. Albany, NY: SUNY Press.
- Miller G (1991). *The science of words*. New York: Scientific American Library.
- Miller G, Beckwith R, Fellbaum C, Gross D & Miller K J (1990). 'Introduction to WordNet: an on-line lexical database.' *International Journal of Lexicography* 3, 235–244.
- Moens M & Steedman M (1988). 'Temporal ontology and temporal reference.' *Computational Linguistics* 14, 15–28.
- Moravcsik J M (1975). 'Aitia as generative factor in Aristotle's philosophy.' *Dialogue* 14, 622–636.
- Moravcsik J M (1990). *Thought and language*. London: Routledge.
- Parsons T (1990). *Events in the semantics of English*. Cambridge, MA: MIT Press.
- Pinker S (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge: MIT Press.
- Pollard C & Sag I (1994). *Head-driven phrase structure grammar*. Chicago University of Chicago Press, Stanford CSLI.
- Pustejovsky J (1991). 'The syntax of event structure.' *Cognition* 41, 47–81.
- Pustejovsky J (1995). *The generative lexicon*. Cambridge: MIT Press.
- Pustejovsky J (2001). 'Type construction and the logic of concepts.' In Bouillon P & Busa F (eds.) *The syntax of word meaning*. Cambridge: Cambridge University Press.
- Pustejovsky J & Boguraev P (1993). Lexical knowledge representation and natural language processing. *Artificial Intelligence* 63, 193–223.
- Pustejovsky J & Busa F (1995). 'Unaccusativity and event composition.' In Bertinetto P M, Binachi V, Higginbotham J & Squartini M (eds.) *Temporal reference: aspect and actionality*. Turin: Rosenberg and Sellier.
- Rappaport Hovav M & Levin B (2001). 'An event structure account of English resultatives.' *Language* 77, 766–797.
- Sanfilippo A (1993). 'LKB encoding of lexical knowledge.' In Briscoe T, de Paiva V & Copestake A (eds.) *Inheritance, defaults, and the Lexicon*. Cambridge: Cambridge University Press.
- Saussure F de (1983). *Course in general linguistics*. Harris R (trans.). (Original work published 1916).
- Stern G (1968). *Meaning and change of meaning. With special reference to the English language*. Bloomington: Indiana University Press (Original work published 1931).
- Tenny C & Pustejovsky J (2000). *Events as grammatical objects*. Chicago: University of Chicago Press.
- Trier J (1931). *Der deutsche Wortschatz im Sinnbezirk des Verstandes: Die Geschichte eines sprachlichen Feldes. Band I*. Heidelberg: Heidelberg.
- Weinreich U (1972). *Explorations in semantic theory*. The Hague, The Netherlands: Mouton.
- Williams E (1981). 'Argument structure and morphology.' *Linguistic Review* 1, 81–114.

Limits of Language

G Priest, University of Melbourne, Melbourne, Victoria, Australia

© 2006 Elsevier Ltd. All rights reserved.

One thing that language (to the sense of which notion we will return in a moment) obviously does well is express things. In particular, it can be used to express information, often of a very complex kind. It might therefore be wondered whether there are limits to this: is there any kind of information that cannot be expressed in language (Alston, 1956)? It is not uncommon to hear it said that there is of course such information: one cannot express the taste of a peach, or the color red. But obviously one *can* express such things. The color red is – what else? – red. What is usually meant by inexpressibility claims, say, about redness, is that there is nothing that can be said in words that will conjure up the mental image of red for someone who has never experienced this before. Maybe so. But to identify meaning with such images, though a natural enough view, is hardly tenable. As the 20th-century Austrian philosopher Ludwig Wittgenstein pointed out in the *Philosophical investigations* (Wittgenstein, 1953), there are many words and phrases that conjure up no such images; and if a person is able to use a word correctly in the company of others, what mental images they are experiencing, if, indeed, any at all, is quite immaterial. Their words are meaningful, and so convey their information, in the usual way. (Private images are irrelevant to public meaning.)

It is sometimes claimed that representations such as pictures, maps, and diagrams can convey information that cannot be captured verbally. No doubt they can often express information more effectively, but it is hard to find examples of information expressible only in this way (especially once one has jettisoned the view that mental imagery has any intrinsic connection with meaning). And, in any case, it is natural enough to think of representations of this kind as languages. It seems profitless (to me, anyway) to dispute over whether or not such things *really* are a language. They share with verbal language at least this: they are structured forms of representation that can be used to convey information of a kind that may never have been expressed before. If, therefore, we are looking for information that cannot be represented, we will have to look elsewhere.

It is presumably uncontentious that relative to most – maybe all – systems of representations there will be information that cannot be expressed. Thus, a medieval monk did not have the conceptual resources

to speak about microchips and quantum fields. It can still, of course, be represented by some other system – maybe the old one augmented by the appropriate concepts. Similarly, relative to any abstract system of representations, there are likely to be things that can be represented, but that, because of their computational complexity, outstrip the resources available to a human brain, and so are inaccessible. These things could become accessible with the help of a different form of representation, however. (Thus, the multiplication of numbers using Roman numerals is computationally much harder than multiplication using Arabic numerals.) The interesting question is whether there is information that is not just unrepresentable because of contingent constraints of this kind but whether there is information that is essentially so. If there is, it is unlikely that this will be demonstrable without appeal to some substantial metaphysical views.

One such view concerns the nature of God. In certain kinds of Christian theology, God is taken to be so different in kind from anything that humans can conceive of that no human concepts can be correctly applied to Him. Because all language deploys only such concepts, the true nature of God cannot be expressed in language (Alston, 1998).

The claim that the nature of God is ineffable is sometimes buttressed by other considerations, especially in the Neo-Platonist tradition. God, it is claimed, is the ground of all beings. That is, He is that which creates and sustains all beings. As such, He is not Himself a being: not a *this* rather than a *that*. His nature cannot, therefore, be communicated in words: to say anything about Him would be to say that He is a *this*, rather than a *that*, and so treat Him simply as another being.

The thought that beings have a ground of this kind is not restricted to Christianity but seems to be a perennial one. It is found in Neo-Platonism (Christian and non-Christian), in which the One plays this role (O'Meara, 1993); it is found in Hinduism, in which Brahman plays this role; it is found in Taoism, in which the Tao plays this role (Lau, 1982); it is found in the writings of the 20th-century German philosopher Martin Heidegger, in which Being (*Sein*) plays this role (Krell, 1977).

A closely related, but different, view is that there is a fundamental or ultimate reality such that the reality that we perceive or conceive is obtained by the imposition of a conceptual grid thereupon. To say what it is like, in itself, is therefore impossible, as anything said about it will deploy our conceptual grid, which is simply a superposition. Again, the existence of such

a reality seems a perennial thought. It is the role played by chora ($\chi\omega\rho\alpha$) in Plato's *Timaeus*; it is the role played by ultimate reality (emptiness, *sūnyatā*) in various branches of Mahayana Buddhism, especially Yogacara (Williams, 1998). Indeed, when Taoism and Indian Mahayana fused to give Chan (Zen), this theme merged with the previous one. Fundamental reality (Buddha nature) can be appreciated only via a direct experience. It is a simple 'thusness' or 'suchness,' beyond all words (Kasulis, 1989).

In some ways, the views of the 18th-century German philosopher Immanuel Kant, as expressed in his *Critique of pure reason* (Kemp Smith, 1923) are similar. For Kant, the empirical world is not independent of us but is partly constituted by our mental concepts. These include the forms of space and time. More particularly for present concerns, these include the logical categories that we apply when we make judgments (such as *all/some, is/is not*). These categories, moreover, depend for their applicability on temporal criteria. Reason forces us, however, to think (about) reality as it is in itself, independent of our mental constructions 'things in themselves' (*dinge an sich*). And because such things are outside time, there is no way that we can apply our categories to them, and so make judgments about them. Thus, although we are forced to recognize the existence of such a reality, there is nothing that can be said about it.

About a century and a half later, but for quite different reasons, Wittgenstein ended up in a similar situation when he wrote the *Tractatus logico-philosophicus* (Pears and McGuinness, 1961). For Wittgenstein, reality is constituted by certain states of affairs; these are composed of objects configured in certain ways. Language, on the other side of the fence, is constituted by propositions; these are composed of names configured in certain ways. A proposition represents a state of affairs if the names in it correspond to the objects in the state, and the configuration of names in the proposition is isomorphic to (has the same form as) the configuration of objects in the state. (This is the so-called picture theory. Wittgenstein is reputed to have been provoked into it by noting how the icons in a scale-representation work.)

It is a consequence of this view that any situation that is not a configuration of objects cannot be expressed in a proposition. Indeed, attempts to do so will produce semantic nonsense. Such situations cannot, therefore, be described. An irony of this is that Wittgenstein's theory itself requires him to talk, not just of objects but also of propositions, configurations, form; and for various reasons these cannot be objects. (For example, propositions can be asserted or denied; objects cannot. And the form of a state of

affairs is not one of the objects in it: it is the way that those objects are structured together.) His own theory is therefore an attempt to do the impossible. This triggers the spectacular dénouement to the *Tractatus*, in which Wittgenstein pronounces his own theory to be nonsense.

The final considerations that I will mention that drive toward things inexpressible concern the infinite, as it is understood in modern logic and mathematics (Moore, 1985). According to this, there are different sizes of infinity. The smallest of these, *countable infinity*, is the size of the natural numbers (0, 1, 2, ...). Because the totality of objects (or even of numbers) is larger than this, so will be the totality of facts about them. (For each object, for example, it is either finite or infinite.) But any language (at least of the kind that is humanly usable) can be shown to have only countably many sentences. There will therefore be many facts that cannot be expressed.

Of course, for all that this shows, each of these facts could be expressed by some richer language; for example, one obtained by adding another name. But there is more to it than this. To say something about an object, one has to refer to it; to do this, one has to be able to single it out in some way; and the totality of all objects is so rich that it will contain objects that are entirely indiscriminable from each other by our finite cognitive resources, and so that cannot be singled out. (Points in a continuum, for example, may be so close as to be indistinguishable by any cognitive mechanism.) There is much, therefore, that will be inexpressible.

As we have now seen, there is a wide variety of metaphysical views that deliver the conclusion that there are things that are ineffable. Evaluating these views goes well beyond anything possible here. I will conclude with a brief discussion of a structural feature of (discussions of) the ineffable.

As is probably clear, theories that claim that certain things are inexpressible have a tendency to say just such things. Thus, Christians say much about God, Buddhists say much about emptiness, Heidegger says much about Being, Kant says much about *dinge an sich*, and Wittgenstein says much about the relation between language and reality. What is one to say about this?

The only thing one can say is that these claims are either literally false or meaningless. The second move is made by Wittgenstein in the *Tractatus*. The first move is more common: all one can do is *deny* any claim made about the object in question. One finds this move in Christian negative theology (Braine, 1998) and some versions of Hinduism, for example, the Advaita Vedānta of Śāṅkara (Sengaku Mayeda, 1992). Kant struggles with a version of this

view when he distinguishes between a legitimate negative notion of *ding an sich* and an illegitimate positive one.

This cannot be the whole story, however, as each position does appear to endorse various claims about the ineffable. How are these to be understood? The most common move is to suggest that one has to understand such assertions as metaphorical, analogical, or in some other nonliteral way. So understood, they can 'point to' the ineffable, although not express it. In Christian theology, this move is made by the 11th-century theologian St. Anselm; similar claims also can be found in the Zen tradition; and Heidegger uses the notion of writing under erasure ('~~x~~') in an attempt to indicate that his words are not to be taken literally.

There is something very unsatisfactory about this, though. One thing that each tradition gives is a set of reasons as to why the thing in question cannot be described: God is beyond categorization; the ground of being is not itself a being; ultimate reality has no features; categories cannot be applied to things outside time; propositions and form are not objects. If one does not understand these claims as literally true, then the very ground for supposing the things in question to be inexpressible falls away. (If 'Juliet is the sun' is not to be taken literally, there is no reason to suppose that she is made of hydrogen and helium.)

Indeed, at the very heart of the view that language has limits is a fundamental paradox (Priest, 1995). To claim that language has limits is to claim that there are things that cannot be talked about; but to say this is exactly to talk about them. The paradox manifests itself in a precise form in some of the paradoxes of self-reference in modern logic. There are many ordinal numbers that cannot be referred to. So there is a least such. But 'the least number that cannot be referred to' refers to that number – Köning's paradox. For skeptics about sizes of infinity, there is even a finite version of this. There is only a finite number of names (i.e., proper names or definite descriptions) with less than (say) 100 letters (in English); there is therefore a finite number of (natural) numbers that can be referred to by names of this kind. So there will be numbers that cannot be referred to in this way. 'The least number that cannot be referred to with less

than 100 letters' (which has less than 100 letters) refers to one of these – Berry's paradox. Various responses to these paradoxes have been proposed in modern logic, but none that is either generally accepted or unproblematic.

One reaction to the fundamental paradox is to reject the notion of the limits of language altogether: there is nothing that it is beyond the ability of language to express. But any theory according to which there are limits to language – including, it would seem, contemporary logic – would appear to be stuck with this contradiction.

See also: Thought and Language: Philosophical Aspects.

Bibliography

- Alston W P (1956). 'Ineffability.' *Philosophical Review* 65, 506–522.
- Alston W P (1998). 'Religious language.' In Craig E (ed.) *Routledge encyclopedia of philosophy*, vol. 8. London: Routledge. 255–260.
- Braine D (1998). 'Negative theology.' In Craig E (ed.) *Routledge encyclopedia of philosophy*, vol. 6. London: Routledge. 759–763.
- Kasulis T P (1989). *Zen action, zen person*. Honolulu: University of Hawaii Press.
- Kemp-Smith N (trans.) (1923). *Immanuel Kant's Critique of pure reason* (2 edn.). London: Macmillan.
- Krell D F (ed.) (1977). *Martin Heidegger: basic writings*. New York: Harper & Row.
- Lau D C (trans.) (1982). *Tao te ching*. Hong Kong: Chinese University Press.
- Moore A (1985). *The infinite*. London: Routledge.
- O'Meara D J (1993). *Plotinus: an introduction to the Enneads*. Oxford: Clarendon Press.
- Pears D F & McGuinness B F (trans.) (1961). *Tractatus logico-philosophicus*. London: Routledge and Kegan Paul.
- Priest G (1995). *Beyond the limits of thought*. Cambridge: Cambridge University Press. 2nd edn., Oxford: Oxford University Press, 2002.
- Sengaku Mayeda (trans.) (1992). *A thousand teachings: the Upadesasahasri of Sankara*. Albany: State University of New York Press.
- Williams P (1998). 'Buddhist Concept of Emptiness.' In Craig E (ed.) *Routledge encyclopedia of philosophy*, vol. 2. London: Routledge. 76–80.
- Wittgenstein L (1953). *Philosophical investigations*. Oxford: Basil Blackwell.

Linguistic Reality

L Wetzel, Georgetown University, Washington, DC, USA

© 2006 Elsevier Ltd. All rights reserved.

Linguistic reality is that portion or aspect of reality that linguistics seeks to understand and explain, and in virtue of which true linguistic claims are true. Linguistics is the scientific study of language, its units, nature, and structure. To see what the units, nature, and structure of language are, we should look to what linguists say. A reasonable place to begin is *An encyclopedia of language*. It reveals that in addition to languages and language users, phonetics cuts up linguistic reality into vowels, consonants, syllables, words, and sound segments, the human vocal tract and its parts (the tongue has five), among other things. Phonology is also concerned with sounds, but organizes them in terms of phonemes, allophones, alternations, utterances, phonological representations, underlying forms, syllables, words, stress-groups, feet, and tone groups. In grammar, morphology describes morphemes, roots, affixes, and so forth, and syntax analyzes sentences, semantic representations, LF representations, among other things. Semantics studies signs, their meanings, their sense relations, propositions, etc. Pragmatics deals with speech acts, speaker meanings, sentence meanings, implicatures, presuppositions, etc. Lexicography investigates nouns, verbs, words, their stems, definitions, forms, pronunciations, origin.

Why Those Objects?

What is the linguist's justification for positing, or recognizing, vowels and consonants, phonemes, morphemes, words, sentences and so on? A general though unhelpful answer is that countenancing such objects produces theories that have a great deal of classificatory, explanatory, and predictive power. The more specific question 'but why posit phonemes?' must be justified within phonology by comparing a theory with phonemes to one without, as must the more specific question of why [b] and [p] are phonemes of English. Why certain noises are enshrined as dictionary entries needs to be justified within lexicographical methodology, rather than, say, within philosophy. The point is that the justification is to be found within linguistic theory, and in most cases within a particular subdiscipline, since subdisciplines have different assumptions. (Phonetics and phonology approach the sound signal differently and hence parse it in terms of different units. Lexicography

might give credence to the thesis that all linguistic objects have instances, since each of the million or so words recognized by the *O.E.D.* have been uttered or inscribed by someone at some time. But grammar might not; there are more, perhaps infinitely many more, sentences than will ever be used.) Assuming there is sufficient internal justification for recognizing the existence of the above-mentioned objects, we may now ask the following question.

What Are They?

What, for example, is that quintessential linguistic object an expression? For that matter, what is a language? One answer to the former would be that an expression is anything that is a noun, verb, adjective, etc., or a sequence of such things built up by means of the grammatical rules. But that merely raises the question of what these other things are. Obviously, nouns and verbs are for the most part **words**, but what are words? Are they physical objects, particular events, abstract objects, kinds, classes, or mental entities? (Parallel questions may be asked of languages: are they social practices, abstract objects, classes of expressions, or psychological products of the language faculty? While it is beyond the scope of this short encyclopedia article to address the more complicated question about language, it is worth noting that the proposed answers mirror those given below on what a word is.)

Types and Tokens

Before such questions can be tackled, we have to disambiguate the word 'word,' which could mean either 'word type' or 'word token.' There is only one definite article in English, the word 'the' – the word type – but there are millions of the's on pages in the Library of Congress and the lips of members of Congress – the word tokens (*see Type versus Token*). Word tokens are particulars; they might be composed of ink, sounds, or smoke signals. They have a unique spatiotemporal location, unlike types. Types are unique but capable of more than one instantiation or representation, unlike their tokens. A good indicator that there are both types and tokens is the different counting procedures associated with the same word. When it is said that an educated person's vocabulary is upward of 10 000 words, that there are exactly 26 letters of the English alphabet, or that English has 18 cardinal vowels, types are being counted, since the numbers would exceed a million if we were counting tokens. As the examples suggest, not only words but

also vowels, phonemes, letters, utterances, sentences, and most of the rest of the linguistic objects found in our survey come in types and tokens.

Which sorts of objects, then, does linguistics explicitly refer to and quantify over – types or tokens? Types – although linguistics is an empirical science based on empirical data, applicable to particulars – tokens – in the causal realm. Yet to ensure generality in an economical format, most of the references and quantifications that appear in linguistic theory involve types and their relationships to each other. Expressions, for example, may be composed of words, and words composed of syllables, and syllables of phonemes, and phonemes of features – all of which are types. (This might be thought puzzling; how can an expression such as ‘Pease porridge hot, pease porridge cold’ be six words long, if there are only four words, each of them unique, of which it might be composed? For a resolution of this puzzle, (see **Type versus Token**). Linguistics, in other words, is awash in references to and quantifications over types, including words. Our original question then becomes the following one.

What Are Word Types?

It should be clear that word types (hereafter: words) are not particular physical objects/events, but it remains to be seen whether they are abstract objects, kinds, classes, or mental entities – or perhaps whether they do not exist at all.

Even among word types, there are several sorts of words. Yet one of the lexicographer’s uses of the word ‘word’ stands out. A rough characterization of this sort is *the sort of thing that merits a dictionary entry*. (*Rough*, because some entries in the dictionary, e.g., *il-*, *-ile*, and *metric system* are not words, and some words, e.g., place names and other proper names, do not get a dictionary entry.) To fix our thoughts on what an account must explain, let us consider the word *color*, or *colour*. According to the *O.E.D.*, the noun *color* is from early modern English, is pronounced [kʌlər] has two ‘modern current or most usual spellings’ [colour, color]; 18 earlier spellings [collor, collour, coloure, colowr colowre, color colure, cooler, couler, coullor, coullour, coolore, coulor, coulure, coulour, culler, cullor, cullour]; and 18 different senses – divided into four branches – with numerous subsenses. The verb *color* is a different word, but with the same pronunciation and spellings (*O.E.D.*, vol. 2: 636–639). *Webster’s* assures us that the word *schedule* has four current pronunciations: [ˈske-(,)jü(ə)l], [ˈske-jəl] (US), [ˈshe-jəl] (Can.) and [ˈshe-(,)dyü(ə)l] (Eng.) (*O.E.D.*, vol. 2: 1044). Thus, a word can be written or spoken; it can have more

than one correct spelling, more than one correct spelling at the same time, more than one sense at the same time, the same correct spelling and pronunciation as a different word; and lastly, a word may have more than one correct pronunciation at a given time. These linguistic facts have to be accommodated by any account of words.

Realism

Probably the most popular account of words is given by platonic realism: words are abstract objects – acausal objects, like numbers, that have no spatio-temporal location. As can be seen from the preceding paragraph, they are very abstract entities indeed, for there is no relatively simple property, like spelling or pronunciation or meaning, that all tokens of the word *color* have in common; not even all their written tokens have the same correct spelling. (Indeed, the realist may argue that that is one of the primary justifications for positing word types – being a token of the word *color* might be the only glue that binds the considerable variety of space-time particulars together). This should discourage the misconception that realism is committed to a platonic ‘form’ (the spelling, say) that all instances resemble the way a cookie resembles a cookie cutter (although the old view lives on in the fact that spellings are called ‘forms’ of the word). Family resemblance in the standard cases is the most that might be hoped for, but intentionality and context are such important factors in determining what type a token is a type of, that even resemblance can fail in nonstandard cases. (It should be noted, if it is not already clear from the foregoing, that a physical object that is a token of a type is not one **intrinsically** – merely by being a certain sequence of shaped ink marks, say. It is only a token relative to a type, a language, and perhaps an orientation. Moreover, it may need to have been produced with a certain intention and in accordance with certain conventions.)

Platonic realism, whether of words or any other abstract objects, must face a serious epistemological challenge, namely, to explain how it is that spatio-temporal creatures such as ourselves can have knowledge of these abstract things if we cannot causally interact with them. Admittedly, we do causally interact with word tokens, but if the tokens are as diverse as was emphasized above, how do we arrive at the properties of the type at all? There are various realist responses to this problem (responses that are not necessarily mutually exclusive). One is to appeal to intuition. Another is to claim that just as maps represent a city, so tokens represent their types, thus reducing the problem to one of characterizing

representation. Another is to reject platonic realism altogether in favor of Aristotelian realism. This involves dropping the claim that words have no spatiotemporal location, and claiming instead that they have many such locations; each type is 'in' each of its tokens. Such a position suggests that there could not be an uninstantiated type. While plausible for words, it is not plausible for sentences. A third response is to claim that words are kinds, just as species are, thus reducing the problem of how we arrive at knowledge of types to one of induction.

Conceptualism

As the terms 'platonic' and 'Aristotelian realism' suggest, we have run into the old philosophical problem of universals in virtue of the fact that types have instances. Not surprisingly, the same camps are in evidence here. The traditional opponents of universals/abstract objects were the conceptualists and the nominalists. The conceptualists argued that there are no general things such as *man*; there are only general *ideas* – that is, ideas that apply to more than one thing. Applied to words, the thesis would be that words are not abstract objects 'out there,' but objects in the mind. Their existence then, would be contingent on having been thought of. While this contingency may have a good deal to recommend it in the case of linguistic items, by itself conceptualism is just a stop-gap measure. For ideas also appear to come in types and tokens (as evidenced by the fact that two people sometimes have the same idea). So either the conceptualist is proposing that word types are idea types – which would be a species of realism – or she is proposing that there are no types, only mental particulars in particular persons, which is a species of nominalism.

Nominalism

The problem for those hostile to universals and abstract objects is to account for our apparent theoretical commitment to types, which are clearly not spatiotemporal particulars. Traditional nominalists argued (as their name implies) that there are no general things, there are only general words, which apply to more than one thing. But this too is not a solution to the current problem, presupposing as it does that there are word types – types are the problem.

Class nominalists have proposed that a word type is just the class, or set, of its tokens. But this is unsatisfactory because, first, classes are abstract objects too, so it is hard to see how this is really a form of nominalism about abstract objects. And second, classes are ill-suited for the job, since classes have

their membership and their cardinality necessarily, but how many tokens a word has is a contingent matter. (One less token would not annihilate the word.)

Initially more promising is the nominalistic claim that talk of types is harmless because it is unnecessary – it is just shorthand for talk of tokens. *The mountain lion is a mammal* is easily translated as 'every mountain lion is a mammal.' So to refer to the noun *color*, say, we need only refer instead to all its tokens. One problem is how to do this. We can't say 'every token of the noun "color"...,' because 'color' refers to a type. And 'every noun "color"...' does not seem grammatical, a fact that is even more apparent if we consider sentences (e.g., 'every "the cat is on the mat"...'). Even if we could, truths will convert to falsehoods (using the 'every'-conversion). The noun *color* is pronounced [kɒ'lər], but particular inscriptions of it are not audible at all. So the question is how we might identify these tokens grammatically but without referring to the noun *color* itself and still say something true and (in some appropriate sense) equivalent. The idea seems to be that the type must embody certain similar features that all and only its tokens have. This is a beguiling idea, until one tries to find such a feature, or features, amid the large variety of its tokens – even the well-formed tokens. Consider *color* and *schedule* again. They demonstrate that neither same spelling, same sense, nor same pronunciation prevail. As the preeminent nominalist Goodman observed, "Similarity, ever ready to solve philosophical problems and overcome obstacles, is a pretender, an impostor, a quack Similarity does not pick out inscriptions that are 'tokens of a common type' Only our addiction to similarity deludes us into accepting similarity as the basis for grouping inscriptions into the several letters, words, and so forth" (Goodman, 1972: 437–438).

Further undermining the reductive approach being considered is that each of the possible defining features mentioned (e.g., spelling, pronunciation) involve reference to types: letter types in the spellings, phoneme types in the pronunciation. (Types are defined in terms of each other.) These too would have to be analyzed away in terms of quantifications over particulars. If, as a last resort, one were to specify a massive disjunction that avoided all references to types, one that captured each and every token of the noun *color*, one would capture tokens of other words, too. Paraphrasing quantifications over word types would be extraordinarily difficult. The moral is that whatever word types are, they are indispensable.

See also: Type versus Token.

Bibliography

- Armstrong D (1986). 'In defense of structural universals.' *Australasian Journal of Philosophy* 64, 85–88.
- Asher N (1993). *Reference to abstract objects in discourse*. The Netherlands: Kluwer.
- Bromberger S (1992). 'Types and tokens in linguistics.' In his *On what we know we don't know*. Chicago: University of Chicago Press.
- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton & Co.
- Collinge N E (ed.) (1990). *An encyclopedia of language*. London: Routledge.
- Goodman N (1951/1977). *Structure of appearance*, 3rd edn. Dordrecht, Holland: Reidel.
- Goodman N (1972). 'Seven strictures on Similarity.' *Problems and projects*. Indianapolis: Bobbs-Merrill.
- Goodman N & Quine W V (1947). 'Steps toward a constructive nominalism.' *Journal of Symbolic Logic* 12, 105–122. Reprinted in Goodman (1972).
- Hale B (1987). *Abstract objects*. Oxford/New York: Basil Blackwell.
- Hutton C (1990). *Abstraction and instance: the type-token relation in linguistic theory*. Oxford: Pergamon Press.
- Katz J J (1981). *Languages and other abstract objects*. Totawa, NJ: Rowman and Littlefield.
- Mish F C *et al.* (eds.) (1993). *Merriam Webster's collegiate dictionary* (10th edn.). Springfield: Merriam Webster, Inc.
- Murray J A H *et al.* (eds.) (1971). *The Oxford English dictionary*. Oxford: Oxford University Press.
- Lewis D (1986a). 'Against structural universals.' *Australasian Journal of Philosophy* 64, 25–46.
- Lewis D (1986b). 'Comment on Armstrong and Forrest.' *Australasian Journal of Philosophy* 64, 92–93.
- Peirce C S (1931–58). *Collected papers of Charles Sanders Peirce*. Hartshorne & Weiss (eds.). Cambridge: Harvard University Press.
- Quine W V (1953). 'On what there is.' In his *From a logical point of view*. Cambridge: Harvard University Press.
- Quine W V (1987). *Quiddities: an intermittently philosophical dictionary*. Cambridge: Harvard University Press. 216–219.
- Simons P (1982). 'Token resistance.' *Analysis* 42(4), 195–203.
- Wetzel L (1993). 'What are occurrences of expressions?' *Journal of Philosophical Logic* 22, 215–220.
- Wetzel L (2000). 'The trouble with nominalism.' *Philosophical Studies* 98(3), 361–370.
- Wetzel L (2002). 'On types and words.' *Journal of Philosophical Research* 27, 239–265.
- Wetzel L (in press). *Types and tokens: an essay on universals*. Cambridge: MIT Press.

Linguistics as a Science

B Clark, Middlesex University, London, UK

© 2006 Elsevier Ltd. All rights reserved.

A common description of linguistics is that it is the 'scientific study of language.' This might seem to be a loose or metaphorical use since the subject matter of linguistics is quite different from what are often thought of as the 'hard' sciences such as physics or chemistry. But linguists are engaged in a process of inquiry that aims to discover facts about the world we live in, and so their work shares important properties of other sciences. Some work in linguistics (e.g., acoustic phonetics) resembles the 'hard' sciences in that it studies physical phenomena in the world. Like psychology, linguistics faces specific issues associated with the fact that its subject matter involves properties of humans, namely, linguistic knowledge and behavior. This article considers some views on what it means to say that a discipline is scientific, what it means to investigate language scientifically, and some different scientific approaches adopted by linguists.

What Is a Science?

The general, popular assumption about what constitutes a science is still probably one based on 'inductivism,' or logical positivism, which was the dominant view of science at the start of the 20th century. In this view, scientists must begin by making sure that they are as objective as possible and simply observe relevant data without prejudging it. They must also make every effort to ensure that they do not themselves affect the data that they are studying. After objectively observing the data, generalizations will emerge and from these generalizations, laws can be derived.

Suppose, for example, that we go to the Antarctic and observe penguins. After a certain number of observations, we might notice that we have never seen a penguin fly. On the other hand, we have seen many penguins swim. So we wonder whether these facts might be the basis of generalizations. We continue to observe and every penguin we observe swims but none of them fly, so we do generalize and come up with the hypotheses that:

- (1) No penguins can fly
- (2) All penguins can swim

These hypotheses can then be tested by further observation. We might also devise specific tests. For example, we might encourage penguins into the air to see whether any of them attempt to fly. Or we might put penguins in water to make sure each of them can swim. Hypotheses (1) and (2) will hold as long as every penguin we observe swims but fails to fly. If repeated tests confirm the hypotheses, then they will be established as verified conclusions.

This approach seems intuitive and clear to most people, but there are serious problems with it. Adapting a diagram from Chalmers (1999: 54), we can represent this way of approaching science as in **Figure 1**.

Perhaps the most fundamental problem with this model is that the derivation of laws and theories is based on induction, which means that they can never be guaranteed to be true. To see this, we need to look at what is involved in induction.

Induction is a process whereby general conclusions are derived based on evidence provided by specific observations in the past. If I go to the same bar often enough and order the same drink, say, a pint of stout, every time, it is possible that the bartender will conclude that I will want a pint of stout every time I come into the bar. If he is keen to make sure I am happy, he might even start pouring my stout as soon as he sees me arrive.

The problem, of course, is that there is nothing to stop me deciding one day to have something different for a change. If I do order a different drink one day, say a glass of lemonade, then the bartender will have wasted his time, and possibly his stout. In other words, conclusions that have been derived through a process of induction are not secure. By contrast, deductive conclusions are guaranteed to be true as long as the premises on which they are based are true. Here are possible steps in the bartender's inductive reasoning process:

- (3a) Billy ordered a pint of stout when he came into the bar on Monday.
- (3b) Billy ordered a pint of stout when he came into the bar on Tuesday.

- (3c) Billy ordered a pint of stout when he came into the bar on Wednesday.
- (3d) Therefore, Billy will order a pint of stout every time he comes into the bar.

And here are two examples of deductive inferences:

- (4a) Billy will order a pint of stout every time he comes into the bar.
- (4b) Therefore, Billy will order a pint of stout when he comes into the bar on Thursday.
- (5a) Billy is drinking stout
- (5b) All stouts are alcoholic drinks.
- (5c) Therefore, Billy is drinking an alcoholic drink.

While (3a–c) could be true and (3d) still turn out to be false, there is no way that (4a) could be true and (4b) false, or that (5a–b) could be true and (5c) false.

The unreliability of inductive inferences is a serious problem for the inductivist approach to science, since it means that no conclusion can ever be safely asserted to be true. You never know for sure that the next penguin you look at won't be able to fly, or whether the next penguin you look at will indeed be able to swim. This means not only that we can't be sure of our conclusions but also that we can never be sure that our scientific endeavors have resulted in any progress. It is always possible that we have just moved from one false hypothesis to another, since we have not been lucky enough to come across the data that would demonstrate our mistake.

Another problem with this view of scientific inquiry is that it is not possible to observe phenomena objectively without first making some assumptions about what might be relevant. Suppose, for example, that I decide I am interested in how children acquire their first language. How will I know which things to observe? Should I begin by observing the speech of the children themselves? Or the speech of other children around them? Or the speech of grownups around them? Or other kinds of behavior exhibited by the children themselves? Or other kinds of behavior exhibited by other children? Or other kinds of behavior exhibited by grownups around them? Or the extent to which they see or hear television? Or their diet? There are countless possibilities and we cannot begin carefully observing data without first guessing which particular data might be relevant. In the same way, hypotheses about whether penguins can fly or swim arise because someone asks the question whether they can. A scientist following the inductivist model might just as easily have decided to look at the penguins' color, their physical behavior, how they mate, how many of them are ever together in one group, and so on. The hypotheses about their status as flightless swimmers will only arise if it is assumed that it might be relevant in this context.

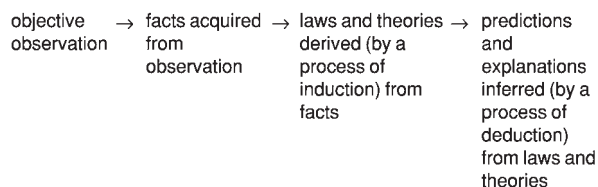


Figure 1 Inductivist model of science (based on Chalmers, 1999: 54).

Not only does the inductivist model of science mean we can never be sure we've made any progress, we also don't have any clear rationale for deciding on our first steps in investigating a particular phenomenon.

These were acknowledged problems for the traditional, inductivist model of science. But no superior model was available until the philosopher Karl Popper (1972, 1979) suggested a new way of thinking about science. He pointed out that, even though we can never be sure an assumption is true, we can demonstrate a conclusion to be false. If I order a glass of lemonade even once, then we know that it's not true that I only ever order stout. If we see just one penguin that can fly, then we know for certain that not all penguins are flightless. He pointed out that falsifying a hypothesis counts as progress, not only because it removes an error from a theory but also because it usually leads to a new, improved hypothesis. Suppose, for example, that we discover a flying penguin. We will not simply reject our initial hypothesis that all penguins are flightless and start again from scratch. Instead, we will wonder what the difference is between those penguins that fly and those that don't. We might, for example, discover that there is a particular penguin species that can fly and come up with a new hypothesis to reflect that.

Another important point Popper made was that the source of hypotheses is much less important than whether they make clear predictions and so can be tested to see if they are false. Newton's theory of gravity, for example, clearly predicts that any object we drop will fall towards the earth. Other things being equal, we know that a floating or rising object will demonstrate that the hypothesis is false. By contrast, if your horoscope claims that 'someone is thinking of you,' it is clearly impossible to show that this is not so. Therefore, this is not a falsifiable, and so not a scientific, claim.

Popper's approach suggests that observation does not need to be the first step in a process of scientific inquiry. Instead, hypotheses can, and indeed must, precede observation. Hypotheses may arise because of objective observation, because of some subjective prejudice, because we dream them, or from any source at all. What is important is that we can test them and so attempt to make progress in our understanding. Popper's view of how science progresses can be represented diagrammatically, as in Figure 2.

The foundations of science, in this view, are not based on objectivity and the notion that observation precedes hypothesis formulation. Instead, what is important is that our hypotheses are clearly formulated and testable. Since we can never verify a hypothesis, we aim instead to develop the best hypotheses we can, and this is evaluated in terms of clarity and falsifiability. Because of the prominence it

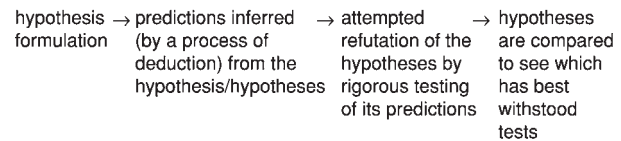


Figure 2 Popperian (falsificationist) model of science.

gives to hypothesis formation and to deductive inferences, Popper's vision of science is also known as the 'hypothetico-deductive' model.

Naturally, there are a number of alternative visions. Imre Lakatos (1970) developed a more sophisticated falsificationist model based on considering the nature of scientific research programs rather than isolated hypotheses, or sets of hypotheses. Thomas Kuhn (1970) suggested a sociological model, in which he described science in terms of paradigms. In Kuhn's view, a scientist proceeds first by learning a current paradigm. He then solves 'puzzles' using the tools available within the paradigm. Sometimes, puzzles are not solvable within the existing paradigm. If this continues, a crisis occurs that eventually leads to a 'revolution' in which the existing paradigm is discarded and a new paradigm takes its place. One important feature of Kuhn's approach is that its sociological nature means that it does not provide a means of distinguishing between scientific and non-scientific statements other than in terms of what is accepted by the relevant group of scientists. A more radical position is taken by Paul Feyerabend (1975, 1978), who claims that the notion of a reliable scientific methodology is an illusion and that scientists have failed to demonstrate that their findings have more value than any other kinds of 'wisdom.'

Whichever philosophy of science is assumed, work in linguistics fits the model as well as any other science. Linguists do not all agree about the nature of their scientific endeavor, but the majority of linguists do see linguistics as a science; and Popper's views on what constitutes scientific activity have been influential in linguistics, just as they have been in other disciplines. (For an introduction to the philosophy of science, see Chalmers, 1999).

The Scientific Study of Language

As Yngve points out, "the origins of linguistic theory can be recognized in Plato and Aristotle, but most clearly in the early Stoics from about 300 to 150 B.C." (Yngve, 1996: 14). Modern scientific linguistics began to develop in the early 19th century in the work of scholars such as Rasmus Rask, Jacob Grimm, Franz Bopp, and Wilhelm von Humboldt. Rask (1830) referred to Linnaeus and Newton in proposing that language was a natural object that should be

studied scientifically. The main focus at this time was on the comparative method, looking at similarities among different languages and using this evidence to reconstruct the ancestors of languages. (For a fuller account of the history of linguistics, see Robins, 1990.) The notion that linguistics is a science has continued since then, while assumptions about what makes linguistics scientific have changed. Perhaps the most significant developments have been the work of Ferdinand de Saussure (1972, 1983), which is usually seen as the starting point of modern linguistics; the development of a rigorous notion of linguistics as a science by Leonard Bloomfield (1926, 1933, 1970) and his American contemporaries in the first half of the 20th century; and the overturning of Bloomfield's approach by the work of Noam Chomsky (1957, 1965). In linguistics today, there are a wide range of approaches, methodologies, and notions of what is scientific, but the Chomskyan approach remains the dominant one.

Saussure's ideas established the notion that linguistics could be 'synchronic' (concerned with a particular language at a particular point in time) as well as 'diachronic' (looking at how a language has developed over time). Saussure's work had considerable influence in the development of structuralist approaches in linguistics and beyond. One particularly significant structuralist approach was that of Bloomfield and his followers in the first half of the 20th century, who developed a much more detailed view of linguistics as a science. As Robins (1990: 208) puts it, "Bloomfield was rigorously scientific, in the light of his own, mechanist, interpretation of science, concentrating on methodology and on formal analysis." It was important for Bloomfield that linguistics should be seen as a scientific enterprise, and his view of what it meant to be scientific was an empiricist one, based on behaviorist psychology. A scientific approach was an empirical study based on objective observation of facts that would lead the scientist to discover facts about languages. It was vital to avoid subjectivity. This meant avoiding hypotheses that did not emerge from objective observation, and it also meant denying the existence of the mind or of mental phenomena. This was because the data alone, considered objectively, did not justify the assumption that mental phenomena really existed. For behaviorists, all behavior could be explained in terms of external stimuli and the reflexes that they caused. Even very complex behavior, such as linguistic behavior, could be understood in terms of complex responses to stimuli. Perhaps the main concern of linguistics was to record facts about particular languages. This was given urgency by the fact that many native American languages were threatened with extinction. It was

important that they should be recorded before they were lost forever. Linguists developed a number of 'discovery procedures' that could be used to scientifically (i.e., objectively) work out facts about the languages being studied.

Chomsky's approach explicitly rejected at least the following assumptions of the Bloomfieldian approach:

- That observation should precede hypotheses
- That the ultimate aim of linguistics was to describe
- That languages were to be understood as collections of utterances (phenomena external to the mind)
- That linguistics should not presuppose the existence of mental phenomena.

Chomsky argued instead that:

- The main aim of linguistics was to construct theories of language and languages
- The ultimate aim of linguistics was to explain language and languages
- That language should be understood as a mental phenomenon (and languages as mental phenomena)
- That there was strong evidence for the existence of the human mind and a cognitive system of knowledge of language.

In general, the key notion in Chomsky's work was that there was convincing evidence for the existence of mental structures underlying human language. This evidence came from the linguistic intuitions of speakers. These included intuitions about what is and is not possible in languages they know. A famous example is the contrast between *eager to please* and *easy to please* illustrated in (6)–(7):

- (6a) John is easy to please
- (6b) It is easy to please John
- (7a) John is eager to please
- (7b) *It is eager to please John

(The asterisk is used here to indicate that most speakers judge (7b) to be unacceptable. It is also used sometimes to indicate the theoretical claim that (7b) is not grammatical in English).

Speakers of English agree that (6a) and (6b) are both acceptable and have similar meanings. Although there is no logical reason for ruling out (7b) and it is easy to see what it would mean by analogy to (6a), speakers agree that (7b) is not an acceptable utterance. This can be explained by the assumption that a cognitive system of knowledge, a grammar, licenses (6a), (6b), and (7a) but rules out (7b).

Using examples like this, Chomsky argued for the existence of mental grammars ('competence') that

underlay actual human linguistic behavior ('performance'). In Chomsky's (1986: 15–51) terms, linguistics should move from the study of 'E-language,' or 'externalized language,' to the study of 'I-language,' or 'internalized language.' This approach revolutionized linguistics, was a major influence on the so-called 'cognitive revolution' that reestablished the mind as a focus of study in psychology, and led to the establishment of the discipline of cognitive science. So what is the scientific methodology of Chomskyan linguistics like? To a large extent, it follows the Popperian model represented in **Figure 2**.

The focus of linguistics is very much on coming up with clearly stated, testable hypotheses, testing them, and constantly updating hypotheses based on how they respond to testing.

However, things are not as straightforward as this. It is not always the case that we have a theory that copes with all of the data until we come across problematic data and we then replace that theory with a new theory. Instead, we compare theories and stick with the theory that we think is the 'best so far' in that it copes best with the relevant data. Most theories do not deal with all of the existing data, and we are usually aware of problems with some data. When this happens, linguists do not reject the theory that cannot cope with the difficult data. Rather, they note the difficulty and continue with their research program until they either find a way to deal with it within the existing theory or until a new theory is formulated that deals with this data and is preferable to the existing theory when looked at overall.

In some cases, data is difficult to interpret. It may be, for example, that speakers are divided over whether a particular form is acceptable. This may correlate with nonlinguistic facts about the group of subjects, such as age (e.g., speakers over a certain age make one judgment while younger speakers make another), class, gender, or geographical location. If so, then the variation in the data can be explained based on these correlations, e.g., as a dialect difference or as a difference in the language of different age groups (which could be an example of a language change in progress). In some cases, though, there may be no obvious correlation with nonlinguistic features of the subjects, i.e., the numbers of speakers who make the different assumptions about the status of a particular form might be comparable in all groups, whether divided by age, class, gender, or geography. When this happens, linguists may 'let the grammar decide'; in other words, they may decide that whatever their existing grammatical theory predicts about the form in question is correct.

In evaluating theories, Chomsky (1986, 2000) has proposed that we can consider three 'levels of

adequacy' that our theories should aim to meet: observational, descriptive, and explanatory adequacy. A theory is observationally adequate if it describes correctly which forms are grammatical. It is descriptively adequate if it also characterizes knowledge that speakers have about those forms. It is explanatorily adequate if it provides an explanation for why the intuitions of speakers are as they are.

It is important to note that, in the Chomskyan view, a linguist's grammar is a theory about the competence of one or more speakers. So to say that a particular expression is 'grammatical' is to make a theoretical claim about the unconscious system of knowledge that speakers have about their language. This is one reason why linguists need to be careful when asking speakers for their judgments on particular utterances. Asking whether a particular form is grammatical may lead to confusion or may generate responses that reflect nonlinguistic (or metalinguistic) assumptions about the relative social status of different forms. So it is often better to ask questions like "Does this sound like a likely utterance to you?" or "Could you imagine someone saying this?"

All sciences involve idealisations. A physicist studying the effects of gravity, for example, wants to observe what a feather and a ball would do if dropped in identical atmospheres and if they were not shaped in such a way that the feather is affected to a greater extent by air resistance than the ball. Similarly, Chomsky points out that the object of study for linguistics is an idealized linguistic system:

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Chomsky, 1965: 3)

This quote has sometimes been misunderstood. Chomsky is not claiming that such a speaker-listener exists. Instead he is saying that linguistics is not concerned with those nonlinguistic properties that lead to different individuals performing differently when making judgments about their language or using it (i.e., with 'performance'). So linguists are studying an object that does not exist as a physical entity. No two speakers share exactly the same language, since all speakers have acquired their language in slightly different circumstances. However, the scientific study of language requires that we abstract away from the differences to find out about the underlying language faculty. This raises a number of methodological issues. Chomsky (1980: 189–192)

makes an analogy between investigation of human language and investigation of thermonuclear reactions taking place inside the sun. It is not possible to set up a laboratory inside the sun and look directly at what is there. Similarly, we cannot view linguistic competence directly. However, we can observe light and heat emanating from the sun and make inferences based on this evidence about what is happening inside the center of the sun. Similarly, we can look at the performance of speakers of a language, including their 'metalinguistic performance' (Birdsong, 1989; Schütze, 1996), when asked to make judgments and make inferences about the system of competence that is involved in determining that performance.

In the Popperian spirit, linguistics should make no assumptions in advance about which data will be relevant in studying language, nor about what methods should be used in studying it. What is important is that hypotheses are clearly stated, testable, and tested. In practice, though, there has also always been a tendency for particular linguists to continue to use the same kinds of methods and data, so that certain groups are referred to in those terms, e.g., as 'corpus linguists' or 'intuitionists.' Much of the work of Chomsky and his followers has been based on evidence from the intuitions of speakers, gathered by personal introspection (i.e., the researcher's own intuitions) interview or questionnaire. But there is no reason in advance to rule out the relevance of other data, whether from corpora, psycholinguistic experiments, or other sources. There are constant debates about the reliability of particular kinds of data and methods for acquiring and interpreting it. Corpus data has been judged problematic because the precise data depends on the accident of who says or writes what at a particular time. Unusual but relevant data is unlikely to appear. The use of intuitions has been questioned because of the risk of subjectivity and the dependence on researchers gathering and interpreting them with enough care. And so on. But the vital thing is that the relevance of particular data is explicitly discussed and justified. As with all sciences, linguists will choose to pay particular attention to some data and to ignore other data based on the hypotheses they have adopted about what will be relevant. But the important question should always be whether the data can be shown to shed light on the particular hypothesis being investigated.

Not all linguists are theoretical linguists. Descriptive linguists aim only to describe languages, not to explain them. A descriptive grammar will aim to make clear what forms exist in a particular language. Any descriptive grammar will have to make idealisations since, as mentioned above, it is not clear that any two speakers will ever have internalized exactly the same system. Where a large percentage

of a language group agree on a particular form, it is fairly easy for the descriptive linguist to write a grammar in agreement with the majority view. In other cases, it may be much harder to decide which option to adopt. The Bloomfieldian American Structuralists who preceded Chomsky were mainly concerned with describing languages. They developed efficient methods for determining facts about languages and provided a wealth of data that theoretical linguists can use in testing their hypotheses. Descriptive linguistics continues to provide useful data with a range of practical as well as theoretical applications.

Not all linguists use falsificationist, or 'hypothetico-deductive,' methods. Conversation analysts, for example, avoid idealisations and use inductive methods to arrive at their conclusions. Conversation analysis has its foundations in ethnomethodology, a branch of sociology that grew out of dissatisfaction with the methods of sociology in the 1960s and 1970s. Research in conversation analysis is qualitative rather than quantitative and avoids early hypothesis formation. One rationale for this is that conversation is complex behavior, and explanations of it will presuppose an understanding of that complexity that cannot be justified until we find out more about the detail of what goes on in conversations. Harvey Sacks (1995), one of the originators of conversation analysis, explicitly states that he views his work as behaviorist and that this approach has much in common with the behaviorist methodology adopted by Bloomfield and the American Structuralists in the early 20th century.

One motivation for using inductivist methods is when research is qualitative rather than quantitative. Lazaraton (2002: 33) suggests that the following features can be seen to distinguish qualitative and quantitative research:

Qualitative Research	Quantitative Research
naturalistic	controlled
observational	experimental
subjective	objective
descriptive	inferential
process-oriented	outcome-oriented
valid	reliable
holistic	particularistic
ungeneralizable	generalizable aggregate analysis
single case analysis	(Lazaraton, 2002: 33)

Qualitative research is appropriate when the aim is to discover the attitudes and experiences of speakers and the motivations and processes behind events, rather than simply counting the number of times a particular event occurs. This is not to suggest that there is a necessary link between particular types of research (e.g., quantitative or qualitative) and

particular models of science (e.g., inductivist or falsificationist); neither quantitative nor qualitative research are necessarily incompatible with any particular model of science.

Not everyone agrees that linguistics is an empirical science. Some reasons for this view stem from the use of intuitions as data. Itkonen (1974), for example, suggested that the reliance on intuitions means that the claims of linguists cannot be falsified. He suggests that the occurrence of an utterance such as **girl the came in* does not falsify the claim that definite articles in English precede nouns in a noun phrase, since this utterance is 'incorrect' and the claim is about 'correct' utterances and sentences. Given this, he suggests that linguistics is different from natural sciences, in which, for example, the discovery of a piece of metal that does not expand when heated would be enough to falsify the claim that all metals expand when heated. Yngve (1996) suggests that the 'ancient semiotic-grammatical foundations' of linguistics are not compatible with modern science. As a result, he suggests putting these foundations aside and replacing them with "new foundations that are fully consonant with modern science as practiced in the more highly developed sciences of physics, chemistry and biology" (Yngve, 1996: 309). Yngve's proposal can be seen as an attempt to reconceptualize linguistics as a 'hard' science. This raises the questions of whether all of the phenomena that have been studied by linguists can fit into this new model and whether the discipline retains its interest if they do not.

For most linguists, however, linguistics is by definition scientific. While much work in linguistics can be understood in some form of Popperian, or post-Popperian, terms, there is nevertheless a wide range of views on the exact nature of the scientific study involved.

See also: Data and Evidence; Language as an Object of Study; Linguistics: Approaches.

Bibliography

- Birdsong D (1989). *Metalinguistic performance and interlinguistic competence*. New York: Springer-Verlag.
- Bloomfield L (1926). 'A set of postulates for the science of language.' *Language* 2, 153–164. [Rpt. in Bloomfield, 1970.]
- Bloomfield L (1933). *Language*. New York: Holt.
- Bloomfield L (1970). *A Leonard Bloomfield anthology*. Hockett C (ed.). Bloomington: Indiana University Press.
- Chalmers A F (1999). *What is this thing called science?* (3rd edn.). Buckingham, UK: Open University Press [1st ed., 1978].

- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky N (1980). *Rules and representations*. Oxford: Basil Blackwell.
- Chomsky N (1986). *Knowledge of language*. New York: Praeger.
- Chomsky N (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Feyerabend P (1975). *Against method: outline of an anarchistic theory of knowledge*. London: New Left Books.
- Feyerabend P (1978). *Science in a free society*. London: New Left Books.
- Itkonen E (1974). 'Linguistics and metascience.' *Studia Philosophica Turkuensia* II. Risteen Kirjapaino, Kokemäki. [Republished as Itkonen E (1978). *Grammatical theory and metascience: a critical investigation into the methodological and philosophical foundations of 'autonomous' linguistics*. Amsterdam: John Benjamins.]
- Kuhn T (1970). *The structure of scientific revolutions*. Chicago: Chicago University Press.
- Lakatos I (1970). 'Falsification and the methodology of scientific research programmes.' In Lakatos I & Musgrave A (eds.) *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press. 91–196. [Reprinted in Lakatos I (1978). Worrall J & Currie G (eds.). *The methodology of scientific research programmes*. Cambridge: Cambridge University Press.]
- Lazaraton A (2002). 'Quantitative and qualitative approaches to discourse analysis.' *Annual Review of Applied Linguistics* 22, 32–51.
- Popper K R (1972). *The logic of scientific discovery*. London: Hutchinson.
- Popper K R (1979). *Objective knowledge*. Oxford: Oxford University Press.
- Rask R ([1830] 1932–1933). 'En Forelæsning over Sprogets Filosofi.' In Hjelmslev L (ed.) *Ausgewählte Abhandlungen*, vol. II. Copenhagen: Levin and Munksgaard. 375–378.
- Robins R H (1990). *A short history of linguistics* (3rd edn.). London: Longman.
- Sacks H (1995). *Lectures on conversation*. Schegloff E (ed.). Oxford: Basil Blackwell.
- Saussure F de (1972). [Originally published 1916.] *Cours de linguistique générale*. Paris: Payot. [ed. Bally C & Sechehaye A, with the assistance of Riedlinger A, with introduction and notes by Mauro T.]
- Saussure F de (1983). *Course in general linguistics*. Harris R (ed.). London: Duckworth.
- Schütze C T (1996). *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Chicago: Chicago University Press.
- Yngve V H (1996). *From grammar to science: new foundations for general linguistics*. Amsterdam: John Benjamins.

Linguistics: Approaches

N Fabb, University of Strathclyde, Glasgow, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Linguistics is the study of language, and there are many different kinds of linguistics, some mutually compatible, some in competition. This diversity of approaches to linguistics is possible because language does not present itself to investigation as a coherent and well-defined field of enquiry that is clearly distinct from other areas of investigation. Instead, language is best imagined as a landscape through which it is possible to take various journeys, its horizons redefined by each approach.

Most approaches to linguistics agree on a few basic facts about language. The first is the fact of linguistic form. While the status of linguistic form is in dispute (as I show below), it is clear that linguistic events such as utterances or inscriptions must be understood as manifestations of linguistic types. Thus utterances must be understood as tokens of combinations of words, even though it may be impossible to isolate a 'word' in the actual stream of sound. The words belonging to classes (such as 'Noun') are made by selecting sounds from an inventory of sounds (phonemes of the language) and so on; in all cases, language must be understood as drawing on inventories of types and combining those types in regular ways: a word of the type 'Article' precedes a word of the type 'Noun' in English (within the Noun Phrase). These regularities, another idealization away from the crude data, involve rules or generalizations or constraints.

No account of language can ignore the fact that language is ordered and regular, based on an inventory of types and rules of combination that together constitute linguistic form. The second fact accepted by all approaches to linguistics is that form relates to meaning. A central function of language is to enable communication, and the organization of linguistic forms has some relation to the organization of meaning. This view – in some form – is shared by many approaches to linguistics: meaning is structured, and the structure of form has some relation to the structure of meaning.

The Status of Linguistic Form

One of the issues that divide approaches to linguistics relates to the 'autonomy' of linguistic form, in a number of senses. One view, expressed from the margins of

linguistics or from outside linguistic theory, is that linguists' discoveries of linguistic form are actually determined by their experience of writing (e.g., the notion of the phoneme is a reimagining of the alphabetic letter), and hence that linguistic form is an artifact of a particular moment in the history of work on language. A second view is that form is entirely determined by function; in its crudest form this is the traditional grammar notion that a word is of the class Noun if it is 'a naming word'; more sophisticated accounts might see the class of a word as fully determined by the function of the word relative to other words in the sentence. A third view sees form as autonomous; thus it is a fact about a word that it is a Noun, this having the same determinate relation to the word as its sound structure or core meaning. While it may have other characteristics as a consequence of being in this word class, including function relative to other words or distinctive phonological characteristics (e.g., stress patterns in English), these functions do not determine its status and are in effect implied by its word class rather than presupposed by it.

Of these three views, the first is an antilinguistic approach to linguistics, the second a strongly functionalist approach, and the third, a strongly formalist approach. Within the formalist approach, one of the issues that arise is which forms are primitives and which are compound or derived forms; for example, some approaches treat Noun as a primitive class, while others see it as derived from features, so that a Noun is a composite of two primitive features, +N and -V (while Verb would be -N +V, Adjective would be -N -V, etc.). The most important early discussion of whether forms are primitives is Halle's argument that there is no coherent notion of the phoneme as a primitive of sound structure; it is instead a composite entity built from the phonological features.

Underlying the discussion of the autonomy of linguistic form is the question of whether any kinds of linguistic forms are specific to language or are also found in other domains. A key text is Chomsky's *Syntactic structures* (1957). He begins his book by discussing generic kinds of forms, as expressed by rules that rewrite a symbol as a string of symbols (phrase-structure rules); simple rules of this kind could, in principle, be found in domains (including cognitive domains) outside language. But he shows that this simple kind of rule is inadequate for understanding language and that a new kind of rule – a transformational rule – must also be used. A transformational rule takes a complex linguistic representation (a tree structure) and changes it into another tree structure, and is specific to language;

transformational rules crucially do not operate on unstructured strings of symbols (e.g., no rule simply inverts the sequence of a string of symbols) but instead operate on more complex structural representations. Later developments in transformational grammar showed that transformations were subject to various constraints (such as the Complex Noun-Phrase Constraint, which prevented a rule simultaneously accessing something inside and outside a complex Noun Phrase); these constraints again were specific to linguistic form. While it has occasionally been argued that transformational rules are found outside language (e.g., in music), it is likely that this kind of rule, which transforms one complex representation into another, is specific to language. These questions are particularly interesting because the existence of some kinds of forms specific to language would support that language is processed in the mind by modes of cognitive organization specific to language.

A basic idea in linguistics is that linguistic form exists to provide formal choices that have functional (semantic or communicative) consequences. For Saussure, a sign such as the word ‘tree’ is a pairing of choices from two distinct systems: (1) the sound system, where the choices produce the three-sound sequence that make up the spoken word and (2) the meaning system, where the meaning of ‘tree’ is chosen from a range of possible meanings. The pairing of sound and meaning is thus a matter of choosing from different inventories. Choice is a dominating principle in Saussurean linguistics and, later, in structuralism, including the nonlinguistic structuralism found in the work of Roland Barthes such as *The fashion system*.

Systemic functional grammar is one of the major approaches to linguistics in which the notion of choice has been fundamental. Each ‘system’ in systemic functional grammar presents ranges of options to choose from when communicating, and the theory investigates the functions expressed by making these choices. This approach has enabled systemic functional grammar to pioneer work in certain areas of linguistic description, some of which (such as Halliday’s early work on thematic roles) have been absorbed into other linguistic theories. In particular, systemicists are interested in the choices among ‘cohesive devices’ in discourse, the devices – conceived of as choices from a range of options presented by the system – explicitly guiding the hearer or reader in understanding how sentences are related.

Systemicists have developed an account of ‘register,’ the set of linguistic features associated with specific genres of verbal behavior (formal interview, informal letter, etc.), where the term ‘register’ itself

is another way of saying ‘range of options from which to choose.’ In addition, systemic functional grammar has a deep interest in stylistic analysis, in which formal choices have functional significance in expressing complexities and nuances of meaning and fit well with ‘close-reading’ approaches in mainstream literary criticism.

Another fundamental idea, shared to some extent by all approaches, is that linguistic form is the mediation between a thought and an utterance; that is, that linguistic form enables the expression of meaning in sound (and writing). Chomsky’s ‘minimalist enquiries’ begin by re-examining this basic idea. From his earliest work onward, Chomsky’s approach to linguistics has always been to ask what the simplest kind of language might be, and then to ask why actual human languages are not this simple.

Thus, as described above, *Syntactic structures* establishes that sentences of a natural language cannot be generated just by rules (phrase-structure rules) of a certain low level of complexity, but require a more complex kind of rule, one that in turn requires a whole theory of linguistic representations which can be subject to transformation.

‘Why aren’t sound and meaning fully aligned?’ is a recently formulated minimalist question; the answer is that organizations of the sentence as meaning are not isomorphic with large-scale organizations of the sentence as sound. In Chomsky’s always-provocative terminology, the question is why language is ‘imperfect.’ Consider, for example, the Noun Phrase that expresses the thing eaten in ‘John ate the cheese’; here this unit of meaning is after the verb, but it is before the verb in ‘the cheese was eaten’ and the word ‘what’ (which substitutes for it, i.e., ‘the cheese’) is at the beginning of the sentence in ‘what did John eat?’ Perhaps the unit of meaning does not always stay in the same place because some other principle, possibly involving the informational structuring of the utterance, forces it to move. Chomsky sees this as an ‘imperfection,’ and seeks to explain why moving the unit is a necessary compromise among the various demands placed on linguistic form by the requirement that speech expresses meaning.

Rationalist and Empiricist Approaches and the Status of Data

A major division in approaches to linguistics pits rationalists against empiricists, as their attitudes towards data exemplify. The key figure for rationalist approaches is Noam Chomsky. Most kinds of linguistics before Chomsky were empiricist approaches (as in the work of Bloomfield), and Chomsky defined his linguistics in opposition to those. Data always

present a problem for linguistics, because it must always be idealized; two slightly different utterances of the same word must be understood by linguistic theory as two tokens of the same type. Without this idealization, there would be nothing to say about language.

Approaches differ in the kinds of idealizations they adopt. Chomsky famously and substantially idealized so that the rules in *Syntactic structures* generate an (infinite) set of sentences corresponding to the set of sentences accepted as grammatical by an idealized user of the language, having separated out all contextual and behavioral factors. The rules describe what someone knows to be grammatical (his or her ‘competence’), not what that person actually says.

Since this rationalist approach to linguistics aims to describe a person’s knowledge of language, not his or her linguistic behavior, there are some significant consequences for the status of data. First, data must be sifted into the relevant and the irrelevant (for this particular purpose); data irrelevant for theoretical syntax might of course be relevant in a different kind of linguistics. It is never clear in advance whether data will be relevant or irrelevant. For example, in *Syntactic structures* Chomsky demonstrates that transformational rules can explain the various configurations of modal, auxiliary, and main verbs in an English sentence; on the other hand, it’s also possible, since a finite and small number of such possible configurations exist, that they could actually just be learned as fixed sequences with no rule involved. There is no guarantee in advance that either approach will be correct.

The second major consequence of a rationalist approach to data is that the frequency of any particular kind of data is ignored, both in the sense that in Chomskyan linguistics ‘variability of rules’ is generally ignored (see discussion below) and in the sense that statistically very rare data may nevertheless have a key theoretical role. Thus, in the early 1980s the rare ‘parasitic gap constructions’ were first discussed; an example is the grammatical sentence ‘Which books did you put on the shelves without reading?’ where the phrase ‘which books’ matches not one but two gaps in the sentence (after ‘put’ and after ‘reading’), one of which is called parasitic. Sentences like this are unlikely to turn up in a corpus, and there are no descriptions of them before this time; nevertheless, such sentences played a major role in helping us understand fundamental aspects of the workings of sentence structure.

The third major consequence for data under a rationalist approach is that these data are invented in order to test the predictions of the theory

(i.e., invented and then tested against speakers’ judgments), rather than gathered and used in constructing a theory. Rationalist approaches to linguistics must assume that the theory will guide the interpretation of the data.

In contrast, empiricist approaches treat theories as constructs that emerge from data. These data may be gathered into a corpus. A fundamental concern is to develop a methodology for gathering, annotating, and understanding the data. Note that rationalist approaches to linguistics do not have this commitment to a methodology, and one of Chomsky’s aims in *Syntactic structures* was to wrest American linguistics from the heavily method-oriented approach associated with the structural linguistics of the 1930s and 1940s.

One advantage of the empiricist approach is its inherent ability, more so than a rationalist approach, to guarantee a result in the sense of a description or annotation of linguistic data. An empiricist approach can be conceptualized as a plan of action, with steps to be followed using clearly specified methods in order to process a collection of data into an account. Because an empiricist approach is supported by agreed methodologies, it can be used to gather and organize large amounts of data, and it is sometimes favored in language work when large amounts of information must be gathered before that information (e.g., from ‘endangered languages’) becomes unavailable.

This kind of empiricist approach also favors the social over the cognitive, because what is gathered is what people say or write, rather than what they know. It has been argued (for example, by Ken Hale) that a rationalist approach, despite its associated risks, is nevertheless best even for endangered languages, because we should be interested in these speakers’ knowledge of their language, not just in their utterances.

While approaches to linguistics agree on the existence of linguistic form, they disagree on how inclusive such a theory of linguistic form should be. One area of disagreement is language statistics. As with any collection of data, linguistic data can be subjected to statistical analysis and this statistical analysis can be described in terms of linguistic rules.

It is a fact that some linguistic rules are ‘variable’ for an individual speaker or for a speech community (the term ‘variable rule’ comes from Labov); an example would be ‘t-glottaling’ in British English or the ‘r-drop’ in some varieties of American English. T-glottaling is the use of a phonetic glottal stop where there is a phonemic /t/ in a word like *bottle* and that can be understood as a consequence of applying a rule changing /t/ to the glottal. Data gathered for a

particular individual will show that the rule is used at some times and not at others, and that these data could be described by attaching a percentage to that rule for that individual (for that data) to indicate how frequently the rule is actually applied in a particular context. The percentage might differ with the social context; for most speakers, t-glottaling is more likely in more informal contexts, and this information could be appended as a statistic to the rule.

The issue dividing approaches to linguistics is whether a way of relating that percentage to the rule can be built into the theory and the representations it permits. The decision depends in part on whether the theory aspires to the holistic or the modular. Holistic theories attempt to incorporate as much explanation of language as possible within a single, connected theoretical model, e.g., Systemic Functional Grammar. Such theories tend to be functionalist and tend not to allow for significant autonomy of form; the holism here claims an explanatory link among various different aspects of language.

Modular theories, on the other hand, are theories that divide language into different subfields – each with its own theoretical account – where there may be limited or no relation among the theories. Modular theories, of which generative grammars are a good example, place considerable emphasis on the autonomy of linguistic form. While generative linguistics on the whole is not interested in statistical facts, Optimality Theory (which has emerged from generative linguistics) is interested; it has had a consistent interest in statistical facts, and various ways have been suggested of incorporating statistical facts into the ranking of constraints.

The distinction between rationalist and empiricist approaches is sometimes entangled in a rhetoric of ‘realism’ or ‘naturalism.’ The terms ‘natural’ or ‘real/realistic’ may be used to valorize certain kinds of (usually empiricist) linguistic theory; hence there is talk of ‘natural phonology,’ ‘psychological realism,’ or ‘real language.’ The underlying assumption in each case is that how linguistics proceeds should be subject to some external constraint on theory-formation, but that is in conflict with the basic principle of rationalist approaches that no theory is guaranteed by some external constraint to be right or successful.

Most commonly, external constraints are drawn from computer science or from psychology, but sometimes a demand is heard that linguistic theory should be constrained by a particular audience’s need to understand it. In each case, these demands are driven by practical anxieties, usually fundamentally financial: Can the theory be used to develop working software? Can it serve psychology? Will the general public understand it sufficiently to want to support it? Will

the linguistics department survive the next round of university cost-cutting?

The Production of Form

All approaches to linguistics agree there is linguistic form. As I have suggested above, these approaches differ on which forms are considered fundamental; for example, are phonemes a fundamental kind of form, or are they a construct based on phonetic features that themselves are the fundamental kinds of forms? They also differ in how linguistic form is generated. What kinds of rules build form? What are the constraints on the operation or output of those rules? Much of the debate in this area has been conducted within kinds of linguistics that refer to themselves as ‘generative’ and trace their intellectual ancestry back to *Syntactic structures*, a book that is a key discussion of the role of rules in language.

In early generative linguistics, there were just rules. These were phrase-structure rules that built a representation called a tree structure, which terminated in a string of words that made up a grammatical sentence, and there were transformational rules that changed one tree structure into another. Within a few years, it was noted that transformational rules were subject to constraints, and these were explored fully by J. R. Ross, who in 1967 wrote one of the most influential Ph.D. dissertations in linguistics, ‘Constraints on Variables in Syntax’ (published as *Infinite Syntax!*, 1986). Ross (1967) showed that the possibility of changing one tree structure into another was prevented if the transformational rule needed to relate two positions, one of which was in an ‘island’ (a relationally defined subarea within a tree structure, such as a sentence inside a noun phrase). This meant that in addition to rules that generated form, there were also constraints on the rules.

The balance between rules and constraints – both in terms of number of rules as opposed to constraints, and complexity of rules as opposed to complexity of constraints – has changed over the course of generative linguistics and is different in different approaches. For example, by the early 1980s, transformational rules were as simple as the rule ‘move something’ (technically formulated as ‘move alpha’), while the constraints on this simple rule were quite extensive and complex.

In addition to constraints on rules, another kind of constraint was introduced: a constraint (or ‘filter’) on the output of the rules. The best known filter in generative syntax of the 1970s was the ‘that-trace’ filter; this stated that a transformational or other rule could not have as its output a sentence in which the subordinating conjunction **that** was followed by

a 'trace' (an empty subject). Thus the sentence 'Who did you say that left?' is ungrammatical because **that** is followed by an empty subject.

Given the possibility of having rules, constraints on rules, and constraints on outputs, some theories have attempted to dispense with one of these elements. In the syntactic theory of the early 1980s called 'government-binding theory,' the terms government and binding did not describe rules but were best understood as expressing constraints on relationships among parts of the tree structure; the role of rules was minimized. In Optimality Theory, both in syntax and in phonology, the role of form-generating rules is minimized, and there are no form-changing rules (such as transformational rules); instead, the burden of explanation is carried by constraints on output.

Optimality Theory, which has been dominant in approaches to sound structure, participates in a key debate about order. In Optimality Theory, constraints are ordered (the term 'ranked' is used) in the sense that an output may violate a constraint and yet be preferred to an output that violates another constraint, because one constraint is ordered above the other. However, this order is fundamentally different from that of the generative phonology that dominated phonology before Optimality Theory, and Optimality Theory now challenges it. In this generative phonology, rules are themselves ordered, such that the input of one rule is determined by the output of the preceding rule. Rule ordering is also the basis of the phonological cycle, where a sequence of rules is applied to a word, then morphological rules are applied (e.g., it is suffixed), and then the same sequence of phonological rules is applied again from the beginning.

The 'Landscape' of Language and its Division into Fields of Linguistic Inquiry

As we have seen, approaches to linguistics can disagree on what aspects of language are open to theoretical description. Alternatively, these approaches may agree on what should be described but dispute the subfield most suited to describing it. Linguistic theory is conventionally divided into distinct but related subfields. For sound structure there is a distinction between phonetics (itself potentially distinguished into acoustic and articulatory phonetics) and phonology; phonetics deals with the mediated manifestations of sound, while phonology deals with the knowledge of (representations of) sound. Here a 'border dispute' involves the distinctness of phonetics and phonology and whether the same kinds of description – e.g., the same kinds of articulatory feature – can be used explanatorily in both phonetics and phonology.

The distinct status of morphology is another area for dispute. In some theoretical frameworks, morphology and phonology are closely intertwined (e.g., lexical phonology and morphology), and thus questions arise about the similarity between morphological and phonological processes. Some morphological processes, such as reduplication (where part of a word is copied to build a larger word), raise difficult problems for phonology. However, morphology is also syntactically relevant, and it is possible that the internal structure of words – the domain of traditional morphology – can in some cases be opened up to syntactic rules. Reorganizing subkinds of linguistics relative to one another can create very productive ways for approaches to linguistics to develop and change.

Another border under dispute is that between the syntax and the pragmatics. This can involve matters of information structure, as well as problematic areas of the syntax such as apposition and conjunction, where there may be pragmatic explanations for apparently syntactic processes. Does the organization of discourse require its own subfield ('discourse analysis') or can it be entirely incorporated under the field of pragmatics? Could a theory of discourse analysis better explain coreference than the syntax?

In his recent work, Chomsky argues that some phenomena previously understood as syntactic are better understood as phonetic – explained as part of the 'phonetic form' of the sentence rather than its syntactic form. Here, for example, an interesting issue is whether the linear order of syntactic elements (such as words) can be entirely explained by some nonsyntactic principle such as the phonetics, on the basis that syntax is the study of hierarchy and phonetics the study of sequence.

Approaches to linguistics can thus differ according to the linguistic material that experts think they can explain under their own subfield. But it is also possible for approaches to different kinds of data to share similar ways of theorizing, a fact that has interesting implications for the organizations of different kinds of linguistic cognition corresponding with these theories. The most significant interplay among domains is seen in generative phonology and generative syntax. At times these have been very close – and at other times very different – in their approaches, raising the question as to whether there are good reasons for thinking that phonology and syntax are fundamentally alike or fundamentally different.

A historical connection, now broken, was in the idea of the 'cycle,' where a linguistic object (such as a word including several suffixes or a sentence including several subordinate clauses) was built in stages; at each stage, the full set of rules was run through

in sequence, starting again from the beginning at the next stage. For a period, it seemed that both words and sentences could be understood as constructed ‘cyclically’ in this way. While the basic notions of this cycle still survive in some forms and in some types of linguistics, few would now argue for the cycle as a point of similarity between the two theories.

Another rich collection of interconnections involves the notion of the ‘feature,’ an idea that also jumped from linguistics to Claude Lévi-Strauss’s anthropology in the 1940s. Form understood as a set of features – each valued as + or – (such that two features can build four derived forms) – has not only been central to generative phonology, but also borrowed as an idea for (1) syntactic categories (e.g., the category of noun is a derived form based on the features +N –V), (2) semantic categories (e.g., ‘agent’), and (3) in lexical semantics (where meanings are considered composites of underlying ‘meaning features’).

More recently, the principles of Optimality Theory, which have been widely used in phonology, have been borrowed to explain syntactic phenomena. Thus, one characteristic of approaches to linguistics is a willingness to borrow ideas from other domains. When the same idea works in different linguistic domains, this

promises to tell us something fundamental about the organization of linguistic form.

Bibliography

- Baker M (1988). *Incorporation: a theory of grammatical function changing*. Chicago: University of Chicago Press.
- Chomsky N (1957). *Syntactic structures*. Berlin: Mouton.
- Chomsky N (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky N (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Halle M (2002). *From memory to speech and back: papers on phonetics and phonology 1954–2002*. Berlin: Mouton.
- Halliday M A K (1985). *An introduction to functional grammar*. London: Edward Arnold.
- Jackendoff R (1977). *X'-syntax: a study of phrase structure*. Cambridge, MA: MIT Press.
- Kager R (1999). *Optimality theory*. Cambridge: Cambridge University Press.
- Kenstowicz M (ed.) (2001). *Ken Hale: a life in language*. Cambridge, MA: MIT Press.
- Labov W (1972). *Sociolinguistic patterns*. Oxford: Blackwell.
- Newmeyer F J (1980). *Linguistic theory in America: the first quarter-century of transformational generative grammar*. New York: Academic Press.
- Ross J R (1986). *Infinite syntax!* Norwood, NJ: Ablex.

Linguistics: Discipline of

N Smith, University College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Language makes us human.

Whatever we do, language is central to our lives, and the use of language underpins the study of every other discipline. Understanding language gives us insight into ourselves and a tool for the investigation of the rest of the universe. Martians and dolphins, bonobos and bees, may be just as intelligent, cute, adept at social organization, and morally worthwhile, but they don’t share our language, they don’t speak ‘human’. Linguistics, the scientific study of language, seeks to describe and explain this human faculty. It is concerned with three things: discovering precisely what it means to ‘know a language’; providing techniques for describing this knowledge; and explaining why our knowledge takes the form it does.

These concerns may seem too obvious to need discussing, but the complexity of our knowledge of language becomes strikingly apparent when we see someone lose their language after they have had a stroke, or when we observe an infant who has yet to acquire the faculty that we deploy so easily. To understand these three concerns, we need a theory, and that is what linguistics provides.

The Meaning of ‘Language’

That linguistics is ‘the scientific study of language’ has become a cliché, but what it means to be ‘scientific’ may not always be obvious, and what people mean when they use the word ‘language’ varies from occasion to occasion. Consideration of what is involved in being scientific is deferred until later in the essay, for now it suffices to observe that only a few aspects of language have been illuminated by theoretical (scientific) linguistics, so there are many areas where it has little, if anything, helpful to say.

The situation is akin to that in biology, viewed as the science of living things. Despite their importance to us, biology has nothing to say about the definition of pets; similarly, despite their relevance to us, linguistics has nothing to say about the definition of dialects.

In everyday usage, 'language' is used differently, depending on whether it is construed as a property of the individual, of society, of the species, or as an autonomous entity in the world. Linguists working in the tradition of 'generative' grammar, the framework that has dominated linguistics for the last 50 years, argue that an 'individual' approach to language is logically prior to any other, but in principle we have the possible domains shown in (1), each suggesting different kinds of questions:

- (1) Language and the Individual
 - Language and the Brain
 - Language and Society
 - Language and the Species
 - Language and Literature
 - Language and the World

Looking at 'Language and the Individual', the central question raised is 'what constitutes our "knowledge of language"?' What properties or attributes does one have to have to be correctly described as a speaker of English, or Burmese, or any other 'natural language' – the term linguists use to refer to languages naturally acquired and spoken by humans, as opposed to the 'artificial' languages of logic or computing? An extension of this question is how and where knowledge of language is represented in the brain, and what mechanisms need to be postulated to enable us to account for our use of this knowledge. Neurolinguistics is an area of remarkable growth, supported by technological advances in imaging.

Under 'Language and Society', sociolinguists raise questions such as 'What are the social variants (class, age, gender, power) that determine, or correlate with, the use of particular forms of the language?' A woman might use some pronunciations or grammatical constructions with statistically significantly greater frequency than a man of the same age, or a female of a different generation. For the world's multilingual majority, social considerations may even determine which language is used in specific situations. A Swiss from Graubünden might use Romansh at home, Swiss German at work, and High German at a conference.

Looking at 'Language and the Species', we might be preoccupied with the puzzle that all human children learn their first language with seeming effortless-ness, while the young of other species, however intelligent, show minimal such ability. Linguists, and researchers in related fields, investigate not only

whether this claim to uniqueness is indeed true but, if it is, how the faculty evolved.

When we turn to the relation between Language and Literature, we confront several issues: 'What is literary form?'; that is, what are the linguistic properties that make something a novel or a novella, a sonnet or an epic? How are literary effects achieved? What are the linguistic characteristics of a particular style?

Looking at 'Language and the World' raises issues of three different kinds. First, how does language relates to things outside the head? That the word 'London' refers to the capital of the United Kingdom is innocuous enough as an informal claim, but it raises interesting, and vexed, philosophical questions. The debate revolves around the status of language as a property of an individual, rather than as an entity with independent existence. This 'external' notion of language is presupposed by those who write irate letters to the press, inveighing against split infinitives, and lamenting the fact that our language is becoming degenerate, either because of the sloppiness of modern youth, the pernicious influence of text messaging, or the role of multiculturalism. The third issue is in many ways the most obvious and the most puzzling: Why are there so many languages? Why does 'human' have so many dialects?

Knowledge of Language

The generativist claim that study of the individual's knowledge of language must be the first or exclusive focus of a scientific linguistics is controversial; that it is a possible, indeed necessary, focus is not seriously in doubt. This individualistic claim implies that linguistics is a branch of psychology, ultimately biology, rather than, say, of sociology. This is not to deny that there are interesting domains of knowledge that take the social conditions of language use as their central focus; it is rather to claim that there is a psychological enterprise which looks at one branch of human cognition and which lends itself to rigorous investigation and, moreover, that it is logically prior to looking at the exploitation of this knowledge in society. This focus on knowledge is highlighted in the claim that the subject of linguistics is 'I-language', rather than 'E-language', where the 'I' stand for *internal* to a particular individual, and 'E' stands for *external* (to the mind of the individual). A corollary of this orientation is that the descriptions that linguists devise are said to be 'psychologically real', where this is not a claim about psychological experimentation or the kind of evidence used in formulating particular linguistic hypotheses, but is simply the claim that we are investigating the human mind and that

current theory is the closest approximation to the truth that we have.

The mind is ultimately a product of the brain (and other systems), and evidence about the mental can sometimes be gleaned from studies of the neural. In general, however, linguists remain agnostic about the details of the relation between the mind and the brain (frequently referring simply to the ‘mind/brain’). That is, we devise theories of a sub-part of human knowledge, but whether that knowledge is localized in the temporal lobe of the left hemisphere, or is distributed throughout the brain, or whatever, is less important. This is not because of lack of interest, but simply because – at present – theories of neural structure are too embryonic to cast much light on linguistic generalizations. Different languages allow different word orders, so that Japanese puts the verb at the end of the sentence and English puts it in the middle. Linguistic theory must provide the means for describing and ultimately explaining this fact, but at present we have no inkling of how the difference between a Japanese and an English speaker might be neurally implemented, so the neurological structure of (this bit of) the language faculty is still a closed book.

What do you have to know to count as a ‘speaker’ of a language? If you say you speak English, it implies that you understand English, too. The point may seem trivial, but knowledge of language is neutral between speaking and hearing; both activities draw on the same fund of knowledge. There is no known illness or accident which leaves you able to speak only English and understand only Portuguese, for instance. This is not to deny that you may be better at talking than listening; or that you may suffer brain damage that leaves you unable to speak while you remain perfectly able to understand. A particularly poignant example of this is provided by Bauby’s autobiographical account of ‘locked-in’ syndrome, where a stroke left the author speechless, but with his language and his ability to understand intact. In normal, non-pathological, cases, however, your ability to *utter* (2a):

(2a) Giraffes have long necks

(2b) Giraffes have necks long

involves the same ability that enables you to *understand* (2a), and also to judge that someone who mistakenly says (2b) has got it wrong. The implication of this observation is that the primary focus of linguistics is on characterizing this neutral knowledge, rather than the mechanisms of speaking, hearing, and judging that are parasitic on it. In other words, linguistics is (largely) about one form of cognition, and only secondarily about the deployment of that

cognitive ability. In the standard terminology, this is known as the ‘competence-performance’ distinction. Your knowledge of language (your competence) underlies your ability to speak, to understand, and to give judgements of well- or ill-formedness (your performance). You may be knocked unconscious and be temporarily unable to speak or understand, but your knowledge typically remains intact – you have competence with no ability for performance. The converse situation, in which you could perform in the absence of any competence, does not occur, though it may characterize the ‘linguistic’ capabilities of the average parrot, which may be able to utter entertaining sentences of what sounds like English, but presumably without the knowledge of English grammar that underlies our abilities.

To count as a speaker of English, you need first to know a large number of words: not just nouns, verbs, and adjectives – words such as *cat* and *go* and *pretty*, whose meaning is relatively transparent, but items such as *the*, *under*, and *however*, whose meaning and use are less easy to specify. Of course, not everyone has the same vocabulary: I may know technical terms in linguistics that you are ignorant of, and you may be familiar with words pertaining to reggae or arachnology that I don’t know, but if either of us were ignorant of words such as *mother* or *and*, people might be justifiably reluctant to classify us as speakers of English.

In addition to knowing the words of a language, you need to know what to do with those words – you need to know the grammar. Our knowledge of language falls into two compartments – the vocabulary (or ‘lexicon’) and the ‘computations’ we can carry out using that vocabulary. This computational system, comprising syntax and morphology, is surprisingly complex, and enables us to produce baroque examples such as Chomsky’s (1995: 88) *Who do you wonder whether John said solved the problem?* Such sentences are of marginal acceptability and citing them may strain the tolerance of outsiders, but this marginal status may itself provide crucial evidence for or against some theoretical claim concerning our knowledge. Henceforth, I shall assume that you and I have the same I-language, abstracting away from differences in vocabulary and grammar. Fortunately, it’s enough for the present purposes to look at the more basic, but nonetheless rich and surprising, knowledge we have of words as simple as *be* and *the*. Consider the examples that follow, which illustrate a wide range of things you know, even if you weren’t previously aware of knowing them. It’s self-evident that *is* and *have* mean different things, as shown in (3), but sometimes they seem to be used interchangeably as in (4):

- (3a) Tom is a problem
- (3b) Tom has a problem
- (4a) Tim is yet to win the Booker prize
- (4b) Tim has yet to win the Booker prize

How is it that something as basic as *is* can sometimes get the same interpretation as *has* and sometimes a different one? Or consider the so-called definite article (*the*), which is often said to mark the distinction between entities which are already familiar and those which are new, as in (5a) and (5b) respectively:

- (5a) My friend likes the penguins
- (5b) My friend likes penguins

But this characterization is not adequate to account for the rather macabre effects found in the newspaper report in (6b) beside the relatively unexceptionable (6a):

- (6a) The woman had lived with the dead man for two years
- (6b) The woman had lived with a dead man for two years

Still less can it account for the fact that on occasion the presence or absence of *the* seems to indicate the difference between subject and object, as in (7):

- (7a) This man is in charge of my brother
- (7b) This man is in the charge of my brother

In (7a) *this man* has control of *my brother*; in (7b) *my brother* has control of *this man*. So what does *the* really mean? Does it even make sense to ask such a question?

Take a more complex example: the word *last* is multiply ambiguous: apart from its use as a noun or a verb, it can function as an adjective meaning either 'final' or 'previous', as illustrated in (8):

- (8a) This is your last chance
- (8b) Your last example surprised me

This ambiguity can result in dialogues which have strikingly different interpretations, as in the alternatives in (9):

- (9) Q "What were you doing in Paris?"
- A1 "Oh, I was collecting material for my last book"
- A2 "Oh, I'm collecting material for my last book"

Answer 1 is itself ambiguous, with either meaning possible for *last* (though 'previous' is the more easily accessible); Answer 2 has only the interpretation that the book under discussion is planned to be the final one I write. The difference must be attributable to the contrast between the past and the present tense,

as that is the only way the sentences differ, but it's not obvious why sentences should be ambiguous or not depending on the tense they contain.

Linguists thrive on such ambiguity, as it regularly provides evidence for structural differences that may not be otherwise apparent. A simple example is provided by the inscrutable notice outside our local school, given in (10):

- (10) This school accepts girls and boys under six

Whether the school accepts girls of any age but only small boys, or no children over six is indeterminate without more information. As we shall see under the section 'Describing Knowledge of Language,' (10) has two quite different syntactic structures corresponding to the two meanings. Similarly the fact that (11) has a number of different interpretations can give us clues as to how to analyze the various possibilities:

- (11) My son has grown another foot

If my son has become taller, the example is parallel to (12a); if he is a freak or a remarkably successful gardener, there are other possibilities, as shown in (12b) and (12c), suggesting that *another foot* in (11) may be correctly analyzed either as a measure phrase or as a direct object:

- (12a) He has grown by another foot
- (12b) He has grown a third foot
- (12c) Another foot has been grown (in this flowerpot)

Such differences of interpretation make the complexity of our knowledge apparent, but unambiguous examples can be just as illuminating and can simultaneously provide evidence against the traditional philosophical claim that meaning can be adequately treated in terms of truth. Thus, we know that (13):

- (13) My first wife gave me this watch

suggests rather strongly that I have been married more than once, but I can utter it truthfully despite having been married only once: my only wife is presumably my first wife. The example is misleading, not false, and so implies that there is much more to meaning than mere truth. As shown by Chomsky's (1957) famous *Colorless green ideas sleep furiously*, structure and meaning (syntax and semantics) can dissociate, so we also know that, despite being initially plausible and syntactically unexceptionable, (14) is meaningless:

- (14) More people have visited Moscow than I have

All the preceding examples illustrate both our knowledge of vocabulary and how it interacts with

(syntactic) structure. The responsibility of linguistics is to describe the full range of such facts, not just for English, but for all human languages. Then, in virtue of its scientific pretensions, it has to (attempt to) explain why these facts rather than any others are the ones that occur – again both in English and in other languages. To do justice to the richness of what we know, it is necessary to distinguish not just the lexicon and the computational system, but to differentiate among syntax, semantics, morphology, phonology and phonetics, and to relate this knowledge to pragmatics – how we interpret utterances in context.

Take our knowledge of morphology, the internal structure of words. We know that *thick*, *thicker*, *thickest*, and *thicken* are all words of English, but that there is no *thinnen* to accompany *thin*, *thinner*, *thinnest*. We know that *thick* relates to *thicken* and that *rich* relates to *enrich*, whereas *richen* is slightly odd, and *enthick* is impossible. This knowledge can't just be a result of our never having heard *thinnen* or *enthick* before, you may never have heard *texted* before, as in "I've just *texted* an urgent message to Fred", but you know that that is possible. As linguists, we may also know that some languages, such as Vietnamese, have almost no morphology: words in this language have none of the internal structure characteristic of affix-rich items such as *indecisiveness* or *rearranged*. On the other hand, some (polysynthetic) languages, such as Inuktitut (Eskimo) or Mohawk pile one affix on top of another so that words are often strikingly complex, and correspond to whole sentences in English. Baker (2001: 87) gives the Mohawk example in (15) with the meaning "He made the thing that one puts on one's body ugly for her":

(15) Washakoty'a'tawitsherahetkvhta'se'

Our knowledge of phonology, the sound structure of language, is equally rich. We know that *past*, *spat*, and *stap* are possible words of English, indeed they all exist; that *stip* and *stup* are also possible words, even though they happen not to exist; but that *satp*, *ptas* and *tpas* are not even possible words. Apart from knowing the segmental make-up of words, we also have knowledge of 'supra-segmentals': that *photograph* is stressed on the first syllable, *photographer* on the second, and *photographic* on the third. Two points need to be made: first, we 'know' this in the sense that we produce the correct pronunciations on demand, and we recognize that deviations from these pronunciations are slips of the tongue or foreigners' mistakes; that is, knowledge of language need not be immediately available to conscious introspection. Second, the characterization in terms of 'first', 'second', and 'third' syllable is actually not the correct

theoretical characterization of our knowledge. As we shall see below, rules of grammar (including phonology) cannot count.

We know more. In an example such as (5a) above, *My friend likes the penguins*, we have to account for the pronunciation of *the* before the initial 'p' of *penguins*: a pronunciation rather different from that of the same lexical item *the* when it occurs before a vowel, as in *My friend likes the otters*. Knowledge of this kind is supplemented by phonetic knowledge which is even harder to bring to consciousness: that the 't' in *photographer* is aspirated, but the 't' in *photograph* is not; that the 'r' in *grime* is voiced, but that in *prime* it is slightly devoiced; that the vowel is longer in *wed* than in *wet*. Such facts belong to the domain of phonetics, the field that deals with the sound properties of language in general, rather than the sound structure of a particular language.

Our phonological knowledge is not self-contained, but may interact in complex ways with our knowledge of the rest of the grammar. We know that (16a) has an alternative pronunciation of the kind given in (16b), where *is* is 'contracted' to's, but that (17a) cannot be matched by the impossible (17b) (impossible is indicated by the asterisk), despite the apparent similarity of the examples:

- (16a) The prime minister is a war criminal
- (16b) The prime minister's a war criminal

- (17a) The president is a war criminal and the prime minister is too
- (17b) *The president is a war criminal and the prime minister's too

An understanding of such asymmetries requires investigation of the relation between syntactic and phonological processes, and relies on an analysis of empty categories: entities that have syntactic and semantic properties but are silent.

In addition to phonology and morphology, we need to account for the (semantic) fact that sentences have meaning. The examples in (18) exploit most of the same words but their meanings are radically different:

- (18a) My friend likes the penguins
- (18b) The penguins like my friend
- (18c) My friend doesn't like the penguins

Moreover, the semantics is 'compositional' – except for idioms, the meaning of a sentence is a function of the meaning of its parts, and their syntactic configuration. The meaning difference between (18a) and (18b) is dependent on which item is subject and which object, notions that can be defined syntactically. In fact, life is a little more complicated than that, as the semantic interpretation of 'subject' is not uniform, and we need to advert to 'thematic relations'

involving ideas of agentivity and patienthood, as shown by the minimal pair in (19):

- (19a) John undertook the surgery reluctantly
(19b) John underwent the surgery reluctantly

John is the subject in both sentences, but is the agent (the surgeon) in the former, and the patient (in both senses) in the latter. These relations are internal to a single sentence, but we also need to relate (the meanings of) different sentences. There are two possibilities: relations which depend on the meaning of individual words, and relations which are purely sentential in that they are independent of such lexical relations. An example of the former is illustrated by (20):

- (20a) Mozart persuaded da Ponte to write a libretto
(20b) Da Ponte intended to write something

where (20b) follows, by virtue of the meaning of *persuade* from (20a). An example of the latter is provided by pairs such as (21a) and (21b), where the truth of (21a) guarantees the truth of (21b):

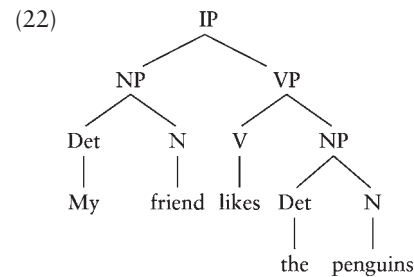
- (21a) Torture is immoral and should be illegal
(21b) Torture is immoral

In the next section, I will outline some of the descriptive mechanisms exploited by (generative) linguistics; then I will try to show how we can approach an explanation for at least some phenomena, looking at a range of examples from English and elsewhere, and use this extension to substantiate the claim that linguistics is a science. Throughout, I shall concentrate on **syntax**. Phonology and phonetics, morphology, and semantics are rich disciplines in their own right, each with a massive literature, but the essence of the analysis of sentences is their **syntactic structure**. And life is finite.

Describing Knowledge of Language

Sentences have structure of various kinds. Returning to the example, *My friend likes the penguins*, we need to describe it in different ways at several distinct ‘levels of representation’: phonological, semantic, and syntactic. Thus, it can be pronounced in a variety of ways – with stress on *friend* or on *penguins*, for instance, with concomitant differences of interpretation. Restricting attention to the syntax, it is intuitively clear that *my* and *friend*, and *the* and *penguins* ‘go together’ in a way that *friend* and *likes*, and *likes* and *the* do not. Each initial word of *My friend* and *the penguins* enables us to pick out some individual or individuals in the world, whereas *friend likes* and *likes the* have no such function within them. This intuition is accounted for in terms of ‘constituency’

represented by means of a simplified tree diagram of the kind in (22):



The top of the tree (IP) indicates that the whole sequence “*My friend likes the penguins*” is an ‘I(nflection) P(hrase)’ (it used to be called ‘Sentence’, but the terminology has changed to reflect changes in our understanding). The IP ‘branches’ into an NP and a VP, where ‘NP’ means ‘Noun Phrase’, that is a sequence consisting of a Noun (N) and something else, and ‘VP’ stands for ‘Verb Phrase’, that is a sequence consisting of a Verb (V) and something else, here in this instance another Noun Phrase. The verb is the (present-tense) form *likes*, and the two Noun Phrases each consist of a Noun (here the singular *friend* and the plural *penguins*) preceded by a ‘Det(eterminer)’, respectively *my* and *the*. Each of ‘IP’, ‘NP’, ‘VP’, ‘N’, etc., are referred to as ‘nodes’ in the tree; IP, NP, and VP, etc., are said to ‘dominate’ everything below them, and to ‘immediately dominate’ everything immediately below them. So VP dominates all of V, NP, Det, N, *the*, and *penguins*, but immediately dominates only V and NP, which are known as ‘sisters’. Once one has got used to the jargon, the advantages of such an analysis are many: it simultaneously shows the linear sequence of items – the order they come in – and the relationships among the component parts: so *the* and *penguins* are more closely related, by virtue of being NPs, than are *likes* and *the* which do not form a ‘constituent’ of any kind. A constituent is defined as any sequence of items that can be traced exhaustively to a single node in the tree: *likes* and *the* can be traced back to VP (and indeed IP), but these nodes dominate other material, too (*penguins*, for instance) so *likes the*, like *friend likes*, are not constituents.

We now have an explicit way of characterizing the example *This school accepts girls and boys under six*. The two interpretations of the object, *girls and boys under six*, can be represented with different constituent structure as in (23):

- (23a) [girls] and [boys under six]
(23b) [girls and boys] [under six]

where the brackets mark the constituents, and indicate that the ‘scope’ of *under six* is respectively either just *boys* (23a) or includes *girls and boys* (23b).

In addition to this syntactic constituent structure, there is morphological structure to deal with: the fact that *penguins* is plural is marked by the addition of the suffix *-s* to the base *penguin*, and the opposite order (with *s-* prefixed to *penguin* to give *spenguin*) is impossible (in English). Investigating the full range of such facts in the world's languages is a matter of intensive research, and addresses the same immediate task of accounting for how it is that native speakers can have the intuitions and make the judgements of well- and ill-formedness that they do.

This last point bears elaborating. One of the surprising facts about our linguistic ability is that it extends to knowing what is impossible as well as what is possible: we have intuitions of ill-formedness or 'negative knowledge'. I have already traded on this fact in assuming that, even though you had probably never heard either example before, you would agree that *Giraffes have necks long* is wrong, whereas *I've just texted an urgent message to Fred* is acceptable. The point can be generalized: the fact that one can recognize mistakes and distinguish them from new but well-formed creations is evidence for the rule-governed nature of the language faculty. Mistakes presuppose norms, or rules. It is also noteworthy that there are 'impossible' mistakes: some logically possible errors just don't happen, even though one might expect them to. Consider an example from language acquisition and the task of the child in working out how questions and statements of the kind in (24) are related:

- (24a) The children are playing truant
(24b) Are the children playing truant?

There are all sorts of hypotheses children might entertain: move the auxiliary (*are*), move the third word, permute the first and second constituents, and so on. The kinds of mistake that children do make, however, show that their hypotheses overlap with these in interesting ways. First, they sometimes make mistakes of a kind for which there is no obvious pattern in the input, even though they may be theoretically well motivated: examples such as the 'auxiliary copying' in (25):

- (25a) Is the steam is hot?
(25b) Are the children are playing truant?

Second, they never try out any hypothesis that would involve them in counting: their attempts always range over modifications of linguistic structure, never of mathematical structure. It seems that all rules in all languages are what is called 'structure-dependent' – they depend on notions such as constituent, Noun Phrase, and so on, but **not** 'third word'. Moreover, children seem not to need to learn this fact – it is a

principle that guides their language acquisition from the start: it is innate. Claims of innateness have been unnecessarily controversial in modern linguistics. No one doubts that humans are innately (genetically) different from cats, chimpanzees, and dolphins, and that this difference underlies our ability to acquire language. Equally, no one doubts that humans acquire different languages depending on the environment they are brought up in: if children are brought up in Turkey rather than Greece, they learn Turkish rather than Greek. It is obvious that both nature and nurture have a crucial role to play. Where controversy is justified, and where empirically different claims can be tested, is in the detail of what needs to be ascribed to the child's 'initial state', of what precisely is innate and what has to be acquired on the basis of experience. Explaining structure-dependence is an area where innateness has been repeatedly (and controversially) defended with a form of argument based on the 'poverty of the stimulus' – the idea that we end up knowing things that it is impossible, or at least implausible, to think that we could find in the input. Consider examples more complex than those above, such as (26):

- (26a) The children who were naughty are playing truant
(26b) Are the children who were naughty playing truant?

If 'moving the third word' or 'moving the (first) auxiliary' were really the correct way of characterizing the relation in (24) one would expect to find example mistakes like that in (27):

- (27a) Who the children were naughty are playing truant?
(27b) Were the children who naughty are playing truant?

Such mistakes simply do not occur. Of course, it is always (usefully) dangerous to say that something does not happen: it may happen in the next utterance one comes across. But this means that the claim is eminently falsifiable (see below), and can anyway be checked by looking for relevant counterexamples in the literature. A nice example of this kind is provided by Neeleman and Weerman's (1997) account of acquisitional differences between Dutch and English. They predicted that Dutch children should, and English children should not, produce sentences with an adverb intervening between a verb and its object, as in (28):

- (28) I will eat quickly the yoghurt

They ransacked the largest international corpus of child data in checking their predictions, and happily found no exceptions.

Formalizing our knowledge of language demands a complex toolkit, only a tiny fraction of which has been given here, but such formalization is a necessary prerequisite to finding explanations, to assimilating linguistics to the scientific enterprise. Given the tools developed here, we can make general hypotheses about the nature of language and begin to test them on a wider range of data from English and elsewhere.

Explanation in Language

Examples involving structure-dependence enable one to address the demand for explanation in addition to description. Let's pursue the issue by looking at the occurrence of items such as *any*, *ever*, or *anything* in English (so-called 'Negative Polarity Items'). At a descriptive level, it is sufficient simply to contrast possible and impossible sentences of the sort seen in (29a) and (29b), where those in (29a) are fully acceptable but those in (29b) are ungrammatical, or infelicitous, or just wrong:

- (29a) John ate something/ some salad
(29b) *John ate anything/ any salad

But *why* is there this contrast? The example in (30) shows that *any(thing)* can occur happily enough in negative statements, but it occurs unhappily in positive statements:

- (30) John didn't eat anything/ any salad

Looking at such negative examples, the generalization seems to be that *any(thing)* needs to occur with (be 'licensed by') a negator. But such an account is inadequate in two different ways: first, (31) shows that it is not just negators that are relevant, but a variety of elements behave in a similar fashion. This class includes questions, conditionals, and other items that there is no space to characterize:

- (31a) Did John eat anything/ any salad?
(31b) If John ate anything/ any salad, I'd be amazed
(31c) Everyone who has any sense has left already
(31d) John denied having eaten any of the cakes

Second, even if we restrict ourselves to negatives, it still seems that life is more complicated than we might wish – (32a) is unsurprisingly fine but, despite being negative, (32b) is unacceptable and none of (32c) to (32e) is acceptable:

- (32a) Something/ some spider bit him in the leg
(32b) *Anything/ any spider didn't bite him in the leg
(32c) *Anything is annoying me
(32d) *Anything isn't annoying me
(32e) *John denied any of the accusations

That is, our first approximation that *any* needs to be licensed by a negative fails in both directions – some

sentences with negatives do not allow *any*; some sentences without a negative do allow *any*. The next obvious assumption might be that *any(thing)* has to be preceded by a negator of some kind (*not* or *n't* here), but (33) shows that this hypothesis is inadequate: it works for (33a) and (33b) but not for (33c) or (33d) – where *nothing* is another negator:

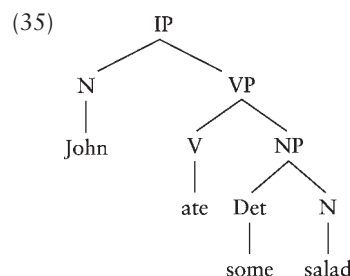
- (33a) The fact that he has come won't change anything
(33b) Nothing will change anything
(33c) *The fact that he hasn't come will change anything
(33d) *That nothing has happened will change anything

The examples in (33a) to (33d) suggest another possibility: perhaps the negator has to be in the same clause as the item (*any*) being licensed? In (33a), the negator and *anything* are in the same clause (compare "*This won't change anything*"), whereas in (33c) and (33d), the negator is in a different clause. We are getting closer, but (34) shows that this is still inadequate as an explanation, as here the negator and *anything* are blatantly in different clauses, but the result is well-formed.

- (34) I don't think he has eaten anything

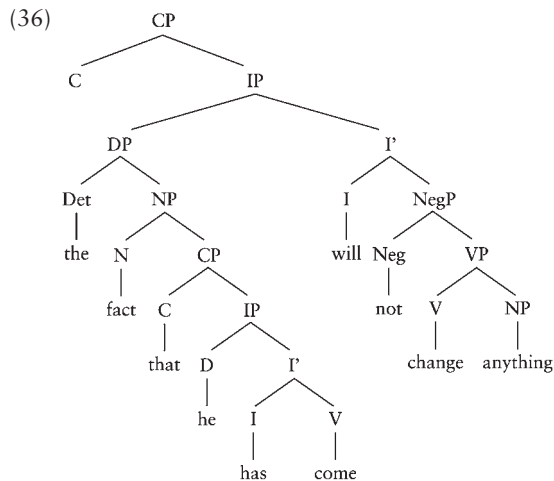
The claim that the negative (or other item) must be in the same clause as *any* fails: some sentences have the negative in a different clause and are nonetheless grammatical; some have the negative in the same clause and are ungrammatical. The correct explanation necessitates an appeal to the notion of 'c-command,' a relation between 'nodes' in a tree. To make this comprehensible and plausible, we need to introduce a little more of the technical machinery of generative grammar.

The representation of sentence structure in terms of trees of the kind shown in (22) can obviously be extended to show the structure of (29a), as shown in (35), where the only novel feature is the uncontroversial claim that *some* is a kind of Determiner:

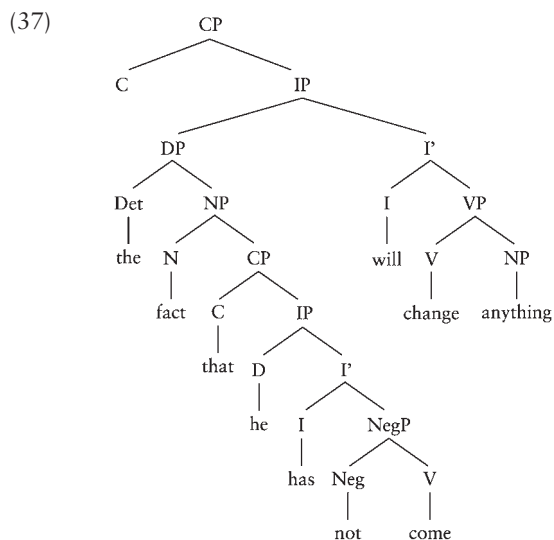


More complex sentences require more complex configurations. Thus, the salient property of an example such as (33a) '*The fact that he has come won't change anything*' is that the subject is not just a noun

or noun phrase, but a noun phrase containing another sentence '*he has come*'. To a first approximation it would have the (simplified) form given in (36), and the ungrammatical example in (33c) **The fact that he hasn't come will change anything* would be characterized by the tree given in (37):



Some of the details of the tree have been included for the sake of those who are already familiar with syntax. So the Complementizer Phrase (CP), optionally headed by a Complementizer such as *that*, and the I' (a constituent intermediate in size between a sentence (IP) and an Inflection element like *will*) are there for the *cognoscenti*. But two things in these trees are important for everyone: first, that they contain a constituent Neg(ation), itself a subpart of a NegP(hrase); and second, that it makes sense to talk of one item being higher in the tree than another. That is, in (36), the 'Neg' is higher in the tree than *anything*, whereas in (37) the 'Neg' is lower in the tree than *anything*.



To make this account rigorous, we need to define exactly what is meant by 'higher' and 'lower', and

that is what is meant by 'c-command': a node A in a tree c-commands another node B if and only if the first branching node dominating A also dominates B. In (36), Neg c-commands *anything* because the first branching node above Neg (i.e., NegP) also dominates the NP *anything*; in (37), Neg does **not** c-command the word *anything* because the first branching node above Neg (again NegP) does **not** dominate *anything*.

It may seem as if we are using a sledgehammer to crack a nut, but the beauty of the analysis is that c-command is not just an arbitrary condition introduced to account for a narrow range of data in English. Rather it extends in two directions: it is a major and essential ingredient in the explanation first of a range of other phenomena in English; and second to a wide range of phenomena in other languages, indeed in all languages: **c-command is universal**.

Before illustrating other uses of c-command, note that if it is universal, we would like an explanation for how that is possible. The obvious answer is that it is innate, part of the faculty of language that differentiates humans from other organisms and explains why all kids but no kittens acquire language. If correct, certain implications follow immediately: c-command is not a condition that children acquiring their first language need to learn, rather (like structure-dependence) it acts as a constraint that determines the kind of hypotheses they can come up with in mastering their first language.

Let us look at one generalization of the usefulness of c-command in English: its use in 'binding theory', the part of linguistics that deals with the distribution of pronouns and reflexives. It is a commonplace that reflexive pronouns such as *myself*, *yourself*, *himself*, and so on, have to agree (or 'be compatible') with their antecedent – the entity they refer back to, so the examples in (38) are fine, but those in (39) are ungrammatical:

- (38a) I admire myself
- (38b) The judge admires himself
- (38c) The waitress might flatter herself
- (39a) *I admire yourself
- (39b) *He admires herself
- (39c) *The waitress flattered ourselves

There are all sorts of other interesting complications with reflexives: if there are two possible antecedents, the sentence is ambiguous, so in (40) *herself* can refer to either *the nurse* or *the woman*:

- (40) The nurse showed the woman some documents about herself

but this is true only if the two potential antecedents are in the same clause as the reflexive: (41)

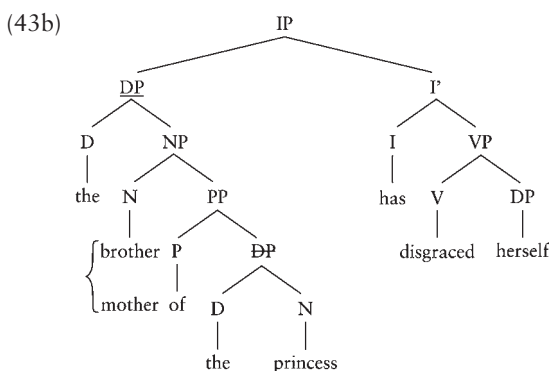
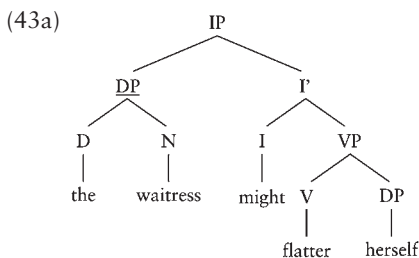
is unambiguous, and *herself* can refer only to *the princess*, because *the queen* is in a different clause:

- (41) The queen said the princess had disgraced
herself

Neither of these extra considerations accounts for why (42a) is unambiguous and (42b) is simply ungrammatical:

- (42a) The mother of the princess has disgraced
herself
(42b) *The brother of the princess has disgraced
herself

The question is why *herself* in (42) can't refer back to *the princess*, but only to the *mother* or the *brother*, resulting in the judgements indicated. The answer is that the antecedent of the reflexive must not only be compatible and in the same clause, but must also c-command it. The structure of possessives such as *the princess's mother* or *the mother of the princess* is a matter of contention, but what is not in dispute is that *princess* is lower in the tree than *mother* or *brother* and hence does not c-command the reflexive: compare the trees in (43a) and (43b) for (38c) and (42a):



In both trees, the underlined DP (*The waitress* in (43a), *The brother/mother of the princess* in (43b)) c-commands *herself*, but the crossed-out DP *The princess* in (43b) does not c-command *herself* so cannot act as its antecedent.

C-command is pervasive in the syntax of English, not just in accounting for polarity items and reflexives. More strikingly, it is pervasive in the syntax of

every other human language. Consider (Cantonese) Chinese. Cantonese has a rich selection of sentence-final particles with a wide range of meanings from conveying a feeling of intimacy to indicating which element in the preceding sequence is the focus. In English, we can indicate this focus by means of stress, giving rise to the kind of difference in (44a) and (44b):

- (44a) John only **watches** football (he doesn't play it)
(44b) John only watches **football** (not cricket)

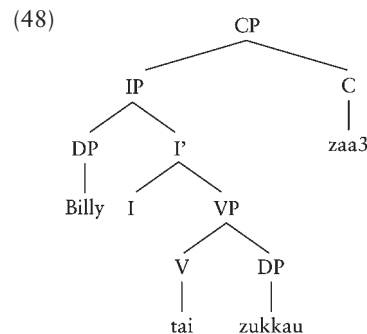
It's even possible, with suitable pause and stress, to have (45b) with the same interpretation as (45a):

- (45a) Only **John** watches football (not Bill)
(45b) **John** only, watches football (not Bill)

Just as in English, Cantonese uses stress to identify the intended focus from the set of possible foci, and the operator *zaa3* (*only*) then associates with this intended focus, as in (46), which can have the various interpretations shown in (47):

- (46) Billy tai zukkau zaa3
Billy watch football zaa3
(47a) Only **Billy** watches football (not Peter)
(47b) Billy only **watches** football (he doesn't play it)
(47c) Billy only watches **football** (not cricket)

There is good evidence (see Law, 2004) that *zaa3* occurs in some C position of the sentence, and hence c-commands everything preceding it in the example in (46): see the tree in (48), (C is arguably final in Cantonese, not initial as it is in English):

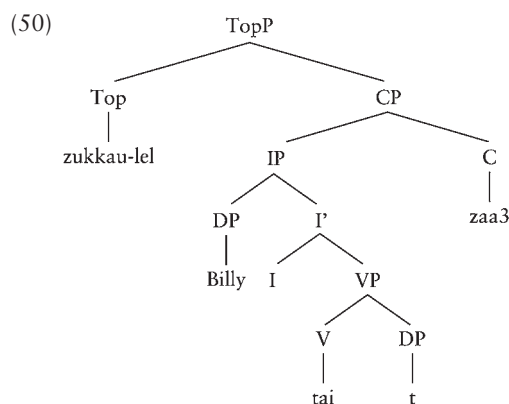


But to talk simply in terms of linear precedence or word order is inadequate. Cantonese also has a process of topicalization whereby a constituent – e.g., *zukkau* ('football') – can be moved to the front of the sentence, where it is attached even higher in the tree than *zaa3*, and marked with *le1* (the 1 indicates a high level tone). This is shown in (49a), with a range of putative translations in (49b) to (49d). Crucially, as indicated by #, (49d) is **not** a possible interpretation of the Cantonese sentence.

- (49a) zukkau-le1, Billy tai t zaa3
(49b) Football, only Billy watches

- (49c) Football, Billy only watches
 (49d) #Only football does Billy watch

Why this should be so is indicated in the tree in (50), where *zukkau* is not c-commanded by *zaa3*: (The ‘t’ for ‘trace’ in (48a) and (49a) indicates where the topicalized constituent *zukkau* moved from).



Because *zaa3* does not c-command *zukkau*, the attempted interpretation in (49d) is simply impossible. The examples are extremely simple, indeed extremely oversimplified, but the moral is clear: the same abstract syntactic condition (c-command) operates in Chinese just as it does in English, and in every other language.

It is worth emphasizing that successful explanations for one class of data are good to the extent that they generalize to phenomena for which they were not devised. C-command was not invented to account for Chinese, but the fact that it automatically accommodates quite subtle data in that language lends support to a theory that incorporates it. The point can be illustrated more widely. Every time one draws a tree of the kind illustrated above, one makes predictions about the well-formedness of a host of other sentences. It is striking that the trees in (36) and (37) exhibit a defining property of human language – its recursive power. That is, the possibility of including one sentence inside another sentence, potentially without limit, gives rise to the infinite expressive power of natural language syntax.

Linguistics as a ‘Science’

Making testable predictions of this kind is one of the hallmarks of science, and we can now elaborate on the claim that linguistics is ‘scientific’. For any discipline to be scientific it must satisfy (at least) the conditions in (51):

- (51a) It must seek **explanation**
 (51b) It must pursue **universals**

- (51c) This will necessarily involve **idealization** which may well give rise to a tension between commonsense and science
 (51d) Most crucially, it will make **falsifiable predictions**

The scientific enterprise is a search for explanatory laws or principles. That is, linguists – like physicists or molecular biologist – seek not only data, but also data that can be used as **evidence** for some theoretical claim. Consider again the analysis of reflexives. Early work in linguistics of the sort best exemplified by the work of Bloomfield (1935) provided detailed exemplification of a wide range of reflexive constructions from a variety of languages, but stopped short of trying to explain their distribution. One of the achievements of generative grammar has been precisely to explain – in terms of ‘binding theory’ – why reflexive pronouns have the distribution they do. To elaborate a little on the discussion given already under the ‘Explanation in Language’ section, the appearance of a reflexive pronoun is determined by principle A of binding theory which says that a reflexive must be ‘bound’ in some domain. As we saw, this means that it must have an antecedent which also meets a number of other conditions. Principle A is in contrast with Principle B, which determines the distribution of ‘ordinary’ pronouns. That is, between them the principles account for the range of facts discussed above as well as for the contrast between *John admires him* and *John admires himself*; why one can construe *John* and *him* as referring to the same person in (52b) but not in (52a), even though the latter seems to include the former as a proper subpart, and a host of other facts:

- (52a) John expects to see him
 (52b) I wonder who John expects to see him

Evidence for – or against – the claims of binding theory, or any part of the theoretical edifice, can be drawn from a wide variety of domains: the distribution of words in sentences; the acquisition of their first language by children, and of second and subsequent languages by both children and adults; the historical change of language over time; the processing of language – be it production or perception – in normal and abnormal circumstances; the problems that arise in pathology, as a result of language disturbance caused by a stroke or a tumor, and so on. In every case, explanation calls for concentration on those data that can provide evidence: the data themselves are trivial until embedded in a theory that can produce testable hypotheses.

A concomitant of this search for explanation is that the generalizations made must carry over in relevant ways to all languages, not just to English or Latin or

Chinese. That is, the quest for laws entails that any science must pursue *universals*, even if that means narrowing the domain of inquiry. This position has two implications: first, that the same principles should apply to Dutch and Hindi and Chinese – so ‘all languages’ is to be construed literally; but second, that the domain of application of these principles may not be superficially obvious. To take the second observation first, it is well-known that there are so-called ‘emphatic’ reflexives, as illustrated in (53a) and (53b), which raise difficulties for any simple analysis of reflexivization in general:

(53a) John himself came

(53b) John came himself

These ‘reflexives’, so labeled because they include the morpheme {*self*}, have somewhat different properties from ‘real’ reflexives: for instance, they don’t have any thematic role, (*came* takes only one argument – you can’t ‘come somebody else’), but simply emphasize the importance of the one role mentioned. On the other hand, they clearly do obey some of the same constraints as ordinary reflexives, as witness the peculiarity of the examples in (54):

(54a) *The boy came herself

(54b) *The boy’s mother himself came

This duality suggests that it might be necessary – as a temporary measure – to limit the domain of binding theory to arguments taking a thematic role, leaving the emphatic examples to be accommodated later after further research. The situation is parallel to the development of a scientific theory of motion. For Aristotle, all motion fell within the ambit of his theory of movement, even the movement of flowers growing. Galileo was able to provide a unified account of terrestrial and heavenly motion by restricting attention to mechanical motion and excluding biological growth. This should not be viewed as a retreat to a position where whatever you say turns out to be true, simply because you have excluded those areas where what you say is false. Rather it is an attempt to define an area where we can begin to understand the complexity of the real world by focusing on phenomena that are comprehensible.

This narrowing is of two kinds: first, one can simply ignore data which fall outside the generalization one is attempting to explain; second, there is scientific idealization – the pretence that things are simpler than they really are. This is justified because such simplification enables one to approach an understanding of the abstract principles which underlie complex phenomena. Such idealization in linguistics was first made explicit in Chomsky’s distinction between competence and performance and his claim

that “linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community.” We all know that real speech communities are not homogeneous, but the force of the idealization is that the heterogeneity that does exist is not a necessary component in an adequate characterization of our knowledge of language or how we come by that knowledge. Consider in this latter respect the simplifying assumption – the striking idealization – that first language acquisition is ‘instantaneous’. It is obvious that children take a considerable time to master the intricacies of their first language. Given how complex the knowledge they end up with is, it may still be justifiable to talk of the surprising speed with which they reach this mastery, but it is **not** by any stretch of the imagination ‘instantaneous’. So what is the force of the assumption? Consider the acquisition of negation.

Most, perhaps all, children go through a stage in which they produce negative sentences with the negative marker (*no* or *not* in English) in peripheral position in the sentence – i.e., first or last – as in (55a) and (55b), heard from two different two-year-olds:

(55a) No computer on

(55b) Computer on no

The context made it clear in each case that the force of the utterance was an order not to turn the computer on. Superficially, it looks as if the two children have different grammatical systems (though they were equally proficient at understanding adult instructions, suggesting that their grammar was more sophisticated than might appear). What is relevant here, however, is the fact that – as far as is known – both children will end up with the same grammatical knowledge of English negation. That is, the different stages they go through in their acquisition of the details of the grammar has no effect on the knowledge they end up with – their adult competence. This claim may, of course, be false. It might turn out that adults who uttered (55a) as children have different grammars from those who uttered (55b) as children. It’s possible, but there is no evidence to that effect, and the idealization to instantaneity is accordingly justified. If one of the things we wish to explain is how humans can progress from a stage in which they are apparently language-less to a stage of adult knowledge, it is advantageous to be able to abstract away from the different paths they may take in acquiring that knowledge. The idealization also simplifies the account of the initial state of the language faculty: what needs to be attributed to the mental make-up of human infants to explain the fact that they do, while infant chimps do not, acquire language.

Idealization of this kind is in turn likely to involve a tension between commonsense and science. The claim of instantaneous language acquisition seems blatantly silly until one considers more carefully what it means. Consider a second example, again from first language acquisition. Children regularly mispronounce the words they are learning, sometimes with surprising results, as in the case of the puzzle *puddle*. When he was about two and a half, my son – like many children – used to pronounce *puddle* as ‘puggle’ ([pʌgəl]). He was perfectly consistent, and used to pronounce words of a comparable kind with the same kind of deformation: so *bottle* became ‘bockle’, *pedal* became ‘peggle’, and so on. The obvious explanation for this behavior was that, for reasons of motor control, he was unable to pronounce *puddle*. But at the same time as he made this mispronunciation, he was also making ‘mistakes’ with words such as *zoo*, pronounced as ‘do’, *lazy*, pronounced as ‘lady’, and so on. The result was striking: although he pronounced *puddle* as ‘puggle’, he consistently pronounced *puzzle* as ‘puddle’ ([pʌdəl]), so the reason for the former ‘mistake’ could clearly not be that he was incapable of the appropriate motor control. He could pronounce ‘puddle’, but only as his version of *puzzle* not for *puddle*. So the commonsense explanation of the phenomenon was wrong. An obvious alternative explanation was that he couldn’t hear the difference, but that hypothesis wasn’t much more plausible either, as his pronunciations of the two words were consistently different, indicating that he must be able to perceive the contrast. So the second ‘obvious’ commonsense explanation was equally problematic. The correct explanation was provided by Marcy Macken, who demonstrated that there was a perceptual problem, but not between *puzzle* and *puddle*, but rather between *puddle* and *puggle*. Of course, *puggle* is not a word of English, so I had failed to observe relevant examples. Words like *riddle* and *wriggle* provide a straightforward minimal pair, but they had not been in my son’s vocabulary. Fortunately, Macken observed that other examples made the case as well as the (missing) minimal pair did. Words such as *pickle* were intermittently pronounced ‘pittle’ ([pitʔl]) suggesting that there was indeed perceptual confusion. The puzzle *puddle* could only be solved when the difference between a variety of other examples was simultaneously taken into account.

I have gone on about this example at such length because it illustrates the beauty of being (potentially) wrong. The most crucial part of the scientific enterprise is that it makes testable (or ‘refutable’ or ‘falsifiable’) predictions. Because my son regularly distinguished *puddle* and *puzzle*, and similar examples, I had claimed explicitly that he had no

perceptual problem. Macken showed that I was wrong and, on the basis of my own data, showed how I was wrong, leading to an improvement in our general understanding of language acquisition, and the language faculty more generally. Such falsifiability is pervasive in linguistics as in all the sciences, and suggests that many, perhaps all, our hypotheses and principles will be in need of revision when we get a better understanding of what is going on. It follows that binding theory, which I have appealed to above, is probably wrong, and will need replacing by some more sophisticated theory in due course. Again this is to be welcomed, though we must simultaneously guard against the danger of ‘naive falsificationism’. There are always going to be contrary data that one’s current theory cannot explain. This is not a reason for simply jettisoning the theory and whatever insights it may provide, but a point of departure for refinement and extension. A clear example is provided by the theory of parametric variation, and the striking revision of his earlier work in Chomsky’s current Minimalist Program (1995).

I have suggested that, like all principles of grammar, binding theory should be universal. But there are problems. Even though (virtually) all languages have reflexives, their distribution is subject to slightly different conditions in different languages. Consider again the contrast between (40), *The nurse showed the woman some documents about herself*, and (41), *The queen said the princess had disgraced herself*, where the former is ambiguous but the latter is unambiguous. The contrast was attributed to the fact that (in English) the antecedent of a reflexive must be in the same clause. So far so good, but if one takes equivalent examples in Chinese, it turns out that the equivalent of (40) is unambiguous, and the equivalent of (41) is ambiguous. The theory would appear to have been refuted: a prediction was made, it was tested, and found to be false. But simply giving up the theory would be defeatist, and it would also mean giving up the explanation for the data it does account for.

The solution is interesting: the universality of binding theory (and likewise for other subtheories of the grammar) is maintained, but some latitude is allowed in the definitions involved – they are ‘parametrized’, as the jargon has it. In this case, all reflexives have to have an antecedent, but language learners have to choose (on the basis of the data they are exposed to) among several other options: whether they are learning a language in which that antecedent has to appear in the same clause or in some other well-defined domain; whether the antecedent has to be a subject or can bear other grammatical relations, and others. In Chinese, the antecedent of a reflexive must be a subject, so (40) is unambiguous; on the other

hand, that antecedent does not have to be in the same clause, so (41) is ambiguous. If you are worried that this is too simple a get-out, an analogy with incest may be helpful: all cultures appear to have an incest taboo forbidding sexual relations between relatives (for instance, fathers and their daughters). The taboo is universal. But how that taboo is instantiated is culture-specific: for example, some groups allow cousins to marry, others do not. The situation with regard to language and language-learning is somewhat more complex than the cultural example, because there are many more choices to be made. The acquisitional task is more complex than it would have been if all languages were exactly like English, but it is not as severe as one might fear. The idea is that the full range of parametric choices in language is available to the child prior to experience – they are in some sense innate – and the child's task reduces to choosing from a set of language structures options it already 'knows.'

Beyond Language: Pragmatics and the Language of Thought

We have looked at a wide range of examples illustrating some of our knowledge of phonology, morphology, semantics, and (mainly) syntax, but we also have knowledge that goes beyond words and sentences. Consider (56a) and (56b): as a remark about Fred, (56a) is fine, with stress on *bats* as indicated by the bold print, but as a reply to the question in (56b) it is anomalous:

- (56a) Fred has written a book about **bats**
 (56b) Who has written a book about bats?

Such discursual knowledge must be distinguished both from syntactic knowledge of the kind that tells us that (57) is ungrammatical:

- (57) Fred has written about **bats** a book

and from real world knowledge of the kind that prompts our scepticism about (58a) and (58b):

- (58a) Bananas have legs
 (58b) Your saucer is being aggressive again

Someone who utters (56a) in response to (56b) probably needs remedial English lessons; someone who utters either of the sentences (58a) or (58b) is either a linguist or in need of psychiatric help.

This brings us into the realm of pragmatics, our interpretation of utterances in context, and to the relation of language to thought. The examples in (58a) and (58b) are felt to be odd not because of our linguistic knowledge – you get the same effect whatever language you translate them into – but

because we know that the world isn't like that. It is our encyclopedic knowledge that tells us this, not knowledge of our language (English). However, when we interpret someone's utterances in some context, we habitually use both our knowledge of English (or whatever other language we are using) and our encyclopedic knowledge. Suppose you hear (3a) *Tom is a problem*. Your knowledge of English vocabulary and grammar provides you with a meaning for the sentence, but it doesn't tell you enough to act. Is your interlocutor looking for sympathy, asking you to do something about it, hoping for a denial? Any or all of these may be what you decide is the case on a particular occasion, but you carry out this construal on the basis of your knowledge of the speaker, of Tom, of your past exchanges with both of them, and so on, indefinitely. The core notion involved is what is 'relevant', an idea that has been made explicit in Relevance Theory, an important extension of linguistics. We are now beyond the language faculty and can hand over responsibility to other disciplines; but one final question needs to be addressed: What is language for?

There are two standard answers: for thought and for communication. Both answers are true, but both need a little hedging. First, we can obviously communicate without using language by means of coughs, sniffs, gestures, and so on. But language is far more subtle than any other system known: conveying specific negative or conditional propositions by means of gestures or sniffing is not obviously possible. Innumerable other creatures have complex communication systems, but none of them, as far as we know, have anything with the recursive power of human syntax (see Sperber and Wilson, 1995; Hauser *et al.*, 2002). Second, the system we use to think with must have a great deal in common with the natural languages we speak, but it is not identical to them. The language of thought can include elements that natural languages cannot – visual images, for instance; and natural languages have properties that would be unnecessary, or even unhelpful, in the language of thought – pronouns, for instance. If I tell you that *she is beautiful*, it's of no use to you storing that in memory as 'she' is beautiful; it has to be stored with a name or some other description replacing *she*. Nonetheless, language has a central role in each of these domains, linking perception and articulation on the one hand to thought processes on the other. This means that the output of our language faculty must be 'legible' to these other systems. Language acts as a code linking representations of sound to representations of meaning. These representations must then be in a form that makes it possible for the sensorimotor apparatus to convert them into pronunciations and

percepts, and for the conceptual system to use them for thinking, especially inference.

So, linguistics provides an account of each of syntax, phonology, morphology, and semantics, and how they relate to each other; pragmatics then tells us how such purely linguistic representations relate to the language of thought – the medium in which we think and carry out inferencing. This relation underlies our ability to interpret the world and the people in it, but the linguistic component is only the first step on the journey. We normally take someone who utters ‘torture is immoral’ to *believe* that torture is immoral, and we expect to be able to predict (at least some of) their actions on the basis of this. But people may lie, and about that linguistics has nothing to say.

See also: Data and Evidence; Principles and Parameters Framework of Generative Grammar; Syntax-Semantics Interface.

Bibliography

- Baker M C (2001a). ‘The natures of nonconfigurationality.’ In Baltin M & Collins C (eds.) *The handbook of contemporary syntactic theory*. Oxford: Blackwell. 407–438.
- Baker M C (2001b). *The atoms of language: the mind’s hidden rules of grammar*. Oxford: Oxford University Press.
- Bauby J-D (1997). *The diving-bell and the butterfly*. London: Fourth Estate.
- Bloomfield L (1935). *Language*. London: Allen & Unwin.
- Carston R (2002). *Thoughts and utterances: the pragmatics of explicit communication*. Oxford: Blackwell.
- Chierchia G & McConnell-Ginet S (2000). *Meaning and grammar: an introduction to semantics*². Cambridge, MA: MIT Press.
- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky N (1994). ‘Chomsky, Noam.’ In Guttenplan S (ed.) *A companion to the philosophy of mind*. Oxford: Blackwell. 153–167.
- Chomsky N (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Fabb N (2002). *Language and literary structure: the linguistic analysis of form in verse and narrative*. Cambridge, UK: Cambridge University Press.
- Fromkin V (ed.) (2000). *Linguistics: an introduction to linguistic theory*. Oxford: Blackwell.
- Fromkin V, Rodman R & Hyams N (2002). *An introduction to language*.
- Gussenhoven C (2002). *Phonology: analysis and theory*. Cambridge, UK: Cambridge University Press.
- Hauser M, Chomsky N & Tecumseh Fitch W (2002). ‘The faculty of language: what is it, who has it, and how did it evolve?’ *Science* 298 (Nov. 22), 1569–1579.
- Heim I & Kratzer A (1998). *Semantics in generative grammar*. Oxford: Blackwell.
- Huddleston R & Pullum G K (2004). *The Cambridge grammar of English*. Cambridge, UK: Cambridge University Press.
- Hudson R A (1990). *English word grammar*. Oxford, UK: Blackwell.
- Hudson R A (1996). *Sociolinguistics* (2nd edn.). Cambridge, UK: Cambridge University Press.
- Jackendoff R (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.
- Law A (2004). ‘Sentence-final focus particles in Cantonese.’ Ph.D. thesis, University College London.
- Macken M (1980). ‘The child’s lexical representation: The Puzzle-Puddle-Pickle evidence.’ *Journal of Linguistics* 16, 1–17.
- McGilvray J (1999). *Chomsky: language, mind, and politics*. Cambridge, UK: Polity Press.
- Neeleman A & Weerman F (1997). ‘L1 and L2 word order acquisition.’ *Language acquisition* 6, 125–170.
- Radford A (2004). *English syntax: an introduction*. Cambridge, UK: Cambridge University Press.
- Smith N V (1973). *The acquisition of phonology: a case study*. Cambridge, UK: Cambridge University Press.
- Smith N V (1989). *The Twitter Machine: reflections on language*. Oxford, UK: Blackwell.
- Smith N V (1999/2004). *Chomsky: ideas and ideals*. Cambridge, UK: Cambridge University Press.
- Smith N V (2002). *Language, bananas and bonobos: linguistic problems, puzzles and polemics*. Oxford: Blackwell.
- Sperber D & Wilson D (1995). *Relevance: communication and cognition*. Oxford: Blackwell.

Logic and Language: Philosophical Aspects

G Callaghan, Wilfrid Laurier University, Waterloo, Ontario, Canada

G Lavers, University of Western Ontario, London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Theories of meaning and methods of linguistic analysis are key items in the agenda of contemporary analytic philosophy. Philosophical interest in language gained substantial impetus from developments in logic that took place in the latter half of the nineteenth century. It was at this time that the early modern conception of logic as an informal ‘art of thinking’ gave way to the contemporary conception of logic as a formally rigorous, symbolic discipline involving, *inter alia*, a mathematically precise approach to deductive inference. The systems of symbolic logic that emerged in the later stages of the nineteenth century were fruitfully applied in the logical regimentation of mathematical theories – analysis and arithmetic in particular – and logical analysis became the cornerstone of a general philosophical methodology for a number of influential figures in the first half of the twentieth century. Though the operative conception of logical analysis did not in every case treat language (or ‘natural language’) as the proper object of investigation, close connections between logic and language were stressed by virtually every proponent of the methodology. Our aim in this entry is to discuss these connections as they appear in the work of some of the eminent precursors and purveyors of the analytic tradition.

The Mathematicization of Logic: Leibniz and Boole

Early modern philosophers were typically antipathetic to the formal approach to logic embodied in Aristotle’s doctrine of the syllogism and its Scholastic variants. Among the major figures of the early modern period, Gottfried Wilhelm Leibniz (1646–1716) is distinguished both for his respect for the Aristotelian tradition in logic and for his general emphasis on the importance of formal methods. Leibniz applauded Aristotle for being “the first to write actually mathematically outside of mathematics” (1696: 465). However, it was Leibniz’s own works, rather than those of Aristotle or of contemporary Aristotelians, that in the period did most to advance the conception of logic as a kind of generalized mathematics.

Leibniz’s logical work consists of a number of manuscripts, unpublished in his lifetime, in which he undertakes the construction of a logical calculus. In virtually all of these works, Leibniz represents judgments and logical laws in a quasi-arithmetical or algebraic notation and he assimilates processes of inference to known methods of calculation with numbers (e.g., by substitution of equals). Leibniz’s motivation for this approach stemmed from his early project of a *lingua characteristica universalis* – a symbolic language geared to the logically perspicuous representation of content in all fields of human knowledge. According to Leibniz, the content of any judgment consists in the composition of the concepts arrayed in the judgment as subject and predicate. A judgment is *true* when the predicate concept is ‘contained in,’ or partially constitutive of, the subject concept. For example, the truth of the judgment that all men are rational consists in the fact that the concept *rational* is contained in the concept *man*, as per the traditional definition of ‘man’ as ‘rational animal.’ (For obvious reasons, Leibniz’s conception of truth posed difficulties when it came to accounting for contingently true judgments, and the task of providing an account of contingency compatible with his conception of truth was one to which Leibniz devoted considerable philosophical attention.) All complex concepts can be parsed as conjunctions of concepts of lower orders of complexity down to the level of simple concepts that cannot be further analyzed. Leibniz’s various schemes for a universal characteristic were predicated on the idea that containment relations among concepts could be made arithmetically tractable given an appropriate assignment of ‘characteristic numbers’ to concepts. For instance, in one such scheme Leibniz proposed that the relationship between complex concepts and their simple constituents be represented in terms of the relationship between whole numbers and their prime factors, thus capturing the unique composition of any complex from its primitive components (1679, 1679/1686). In this and similar ways, Leibniz sought to provide a basis for the view that inference, and the evaluation of truth more generally, could be carried out algorithmically – that is, as a mere process of calculation – by familiar arithmetical means.

By the 1680s Leibniz had grown pessimistic about the prospect of completing the project of the universal characteristic, and he turned his energies to the more confined task of devising an abstract logical calculus. Leibniz worked on a number of different versions of his logical calculus through the 1680s and 1690s. In each case he explained how standard propositional

forms could be expressed in a quasi-algebraic notation. He also laid down logically primitive laws pertaining to his formulas (what he called ‘propositions true in themselves’) and abstractly specified valid inferential transformations, usually in terms of a definition of ‘sameness’ in conjunction with principles pertaining to the substitutability of identicals. Though Leibniz’s efforts at constructing a logical calculus were hampered by his view that all judgments ultimately reduce to a simple subject-predicate form – thus excluding primitive relational judgments – his emphasis on formal explicitness and mathematically exact symbolism stands as an anticipation of the main lines of development in formal logic in the following centuries.

Leibniz’s mathematical approach to logic made little impression until the late nineteenth century, when his manuscripts were finally collected and published. By that time, however, the mathematical approach had gained momentum independently, largely on the basis of the work of George Boole (1815–1864). Boole shared with Leibniz the aim of devising an algebraic means of expressing relationships among terms figuring in propositions. However, Boole differed from Leibniz in treating the extensions of concepts (or classes), rather than concepts construed as attributes or ‘intensions,’ as the relevant propositional constituents. In *The laws of thought* (1854), Boole presented his class logic, which he called ‘the logic of primary propositions,’ as the first and fundamental division of his system. In the second part of the same work, Boole adapted the calculus of classes to a special interpretation that allows for the representation of logically compound propositions, or ‘secondary propositions,’ thereby unifying (after a fashion) the calculus of classes with a version of modern propositional calculus.

Boole’s central idea is that an algebra of logic arises as an *interpretive variant* of standard numerical algebra when the latter is modified by a single principle that is naturally suggested by the logical interpretation of the symbolism. In Boole’s class logic, letters (or ‘literal symbols’) are interpreted as standing for classes of things determined by some common attribute, with ‘1’ standing for the universe class and ‘0’ standing for the null class. Multiplication, addition, and subtraction operators are treated as standing for the operations of intersection, disjoint union, and difference (or ‘exception’) of classes, respectively. Primary propositions are then expressed as equations with appropriately formed class terms standing on either side of the identity sign. On the basis of this class-theoretic interpretation of the symbolism, Boole maintained that the logical calculus differs from ordinary numerical algebra only with

respect to the characteristically *logical* law that, for any class x ,

$$xx = x \text{ (} x \text{ intersect } x \text{ is } x \text{)}$$

which holds generally for class theoretic intersection but which holds for numerical multiplication only for $x = 0$ and $x = 1$. Having emphasized this difference, Boole observed that the laws and transformations of numerical algebra will be identical to those of an algebra of logic when the numerical values of literal symbols in the former are restricted to 0 and 1. Boole appealed to this formal analogy between the numerical and logical algebras in justifying his approach to inference, which he presented as a process of solving sets of simultaneous equations for unknowns by standard algebraic methods.

In *The laws of thought*, Boole transformed the calculus classes into a serviceable propositional calculus by interpreting his literal symbols over ‘portions of time’ during which elementary propositions are true, thus adapting the notation and methods designed for dealing with class relationships to the propositional case. Boole’s appeal to portions of time reflected a somewhat puzzling endeavor to assimilate or reduce propositional logic to the kind of term logic embodied in his class calculus, and the artificiality of this approach was not lost on subsequent logicians both within and without the algebraic tradition. However, peculiarities of interpretation notwithstanding, Boole can be credited with the first systematic formulation of propositional logic and a commensurate expansion of the scope of formal logic in general. Moreover, his suggestion that propositional logic, class logic, and numerical algebra (suitably restricted) arise as interpretive variants of a single algebraic system anticipates subsequent developments in abstract algebra and (perhaps only dimly) modern model-theoretic methods in logic.

The contributions of Leibniz and Boole constitute beginnings in the fruitful deployment of artificial languages in the analysis of propositional content and the systematization of deductive inference. However, despite their considerable accomplishments, neither Leibniz nor Boole can be credited with bringing logic to its current state of maturity. They produced no inroads in the logic of relations and the use of quantifiers for the expression of generality is entirely foreign to their work. These shortcomings were addressed by later logicians working in the algebraic tradition (e.g., Pierce and Schröder), but the significance of their resolution for the development of modern formal logic and its philosophical offshoots will be better appreciated if we adopt a somewhat different perspective on the historical interplay between logic and mathematics.

Logic and Language in Frege

For the better part of his career, the philosopher-mathematician Gottlob Frege (1848–1925) devoted his energies to establishing the ‘logician’ thesis that arithmetical truth and reasoning are founded upon purely logical principles. At an early stage in his efforts, Frege realized that existing systems of logic were inadequate for carrying out the analysis of content necessary for establishing arithmetic’s logical character. His *Begriffsschrift* (or ‘concept notation’) (1879) was intended to address this deficiency. The logical system that Frege presented in the *Begriffsschrift* and later refined in Part I of his *Grundgesetze Der Arithmetik* (1893) constitutes the greatest single contribution in formal logic since the time of Aristotle. The most distinctive aspects of Frege’s logic are (1) the use of variables and quantifiers in the expression of generality; (2) the assimilation of predicates and relational expressions to mathematical expressions for functions; (3) the incorporation of both propositional logic and the logic of relations within (second-order) quantificational logic; (4) the notion of a formal system – i.e., of a system comprising a syntactically rigid language along with explicit axioms and inference rules that together determine what is to count as a *proof* in the system.

Frege’s approach to generality is based on his analysis of predication in terms of function and argument. In arithmetic, a term such as $7 + 5$ can be viewed dividing into function and argument in different ways. For instance, it can be treated as dividing into the function $() + 5$ with 7 as argument, or as dividing into the function $7 + ()$ with 5 as argument, or as dividing into the binary function $() + []$ with 7 and 5 (in that order) as arguments. Frege’s approach to predication assimilates the analysis of sentences to this feature of the analysis of arithmetical expressions. For example, a simple sentence such as ‘John loves Mary’ can be regarded as predicating the (linguistic) function ‘ $()$ loves Mary’ of the singular term ‘John,’ or the function ‘John loves $()$ ’ of the singular term ‘Mary,’ or the relational function ‘ $()$ loves $[]$ ’ of ‘John’ and ‘Mary’ (in that order). In the *Begriffsschrift*, Frege remarked that, for simple sentences like this, the analysis into function and argument makes no difference to the ‘conceptual content’ that the sentence expresses. However, the *possibility* of analyzing a sentence in these ways is nevertheless crucial to logic, since only on this basis do we recognize logical relationships between generalizations and their instances. Adopting a standard arithmetical practice, Frege makes use of variables as a means of expressing generality. For example, by putting the variable ‘ x ’ in the argument-place of ‘Mary’ in our example, we

arrive at the statement ‘John loves x ,’ which is the *Begriffsschrift* equivalent of the colloquial generalization ‘John loves all things.’ The inference from this generalization to ‘John loves Mary’ now *requires* that we regard ‘Mary’ as argument to the function ‘John loves $()$,’ since only so is ‘John loves Mary’ recognizable as an *instance* of the generalization. Other function-argument analyses become salient in connection with other generalizations to which the statement relates as an instance (e.g., ‘ x loves Mary’).

In the system of the *Begriffsschrift*, the above described use of variables suffices to express generality in a limited variety of sentential contexts. However, Frege’s broader treatment of generality involves a second crucial component, namely, the use of quantifiers – i.e., the variable binding operators ‘ $\forall x$ ’ (read: ‘Every x ’) and ‘ $\exists x$ ’ (read: ‘some x ’) – as a means of indicating the *scope* of the generality associated with a variable. (Our discussion here prescind from the peculiarities of Frege’s now obsolete notation as well as his convention of treating existential quantification in terms of universal quantification and negation – i.e., his treatment of ‘ $\exists x \dots$ ’ as ‘ $\sim \forall x \sim \dots$ ’). One of the many ways in which the use of quantifiers has proven important to logic concerns the expression of multiply general statements, for which no adequate treatment existed prior to the *Begriffsschrift*. Consider, for example, the relational generalization ‘Everyone loves someone.’ This statement is ambiguous between the following two readings: (1) ‘There is some (at least one) person that is loved by all,’ and (2) ‘Every person is such as to love some (at least one) person.’ The use of quantifiers resolves this ambiguity by requiring that expressions of generality in multiply general statements be ordered so as to reflect scope. The first reading of the statement is expressed by the existentially quantified sentence

$$\exists y \forall x xLy$$

where the scope of universal quantifier falls *within* that of the existential quantifier. (For convenience, we assume here that the variables ‘ x ’ and ‘ y ’ are restricted to a domain of persons.) By contrast, the second reading is given by the sentence

$$\forall x \exists y xLy$$

where the universal quantifier has wide scope with respect to the existential quantifier. Since the *Begriffsschrift*’s formation rules ensure that the scope of a quantifier will be properly reflected in any sentence in which it occurs, an ambiguous sentence such as the one we started with cannot even be *formulated* in the language. Scope considerations apply in essentially the same way to sentential operators (e.g., the negation sign ‘ \sim ,’ and the conditional

sign ‘ \rightarrow ’) in the logic of the *Begriffsschrift*. For instance, the ambiguity of the sentence ‘Every dog is not vicious’ results from the fact that, as stated, the sentence does not determine the scope of the negation sign with respect to that of the universal quantifier. On one reading, the scope of the negation sign falls within that of quantifier, i.e., ‘ $\forall x(Dx \rightarrow \sim Vx)$,’ the statement thus affirming that anything that is a dog is not vicious (or, colloquially expressed: ‘No dogs are vicious’). By contrast, when the statement is read as giving wide scope to the negation sign, i.e., ‘ $\sim \forall x(Dx \rightarrow Vx)$,’ it becomes the denial of the generalization that all dogs are vicious (i.e., ‘It is not the case that all dogs are vicious’). As the above examples begin to suggest, Frege’s technique of ordering of operators according to scope provides the basis for his incorporation of both propositional logic and the logic of relations within quantificational logic.

Frege’s philosophical interest in language extended beyond his characterization the formal mechanisms of a logically perspicuous language such as the *Begriffsschrift*. In the classic paper ‘On Sense and Reference’ (1892), Frege presented a framework for a general theory of meaning applicable to both natural languages and formal languages. The core of the doctrine consists in the contention that any adequate account of the meaning of a linguistic expression must recognize two distinct, but related, semantic components. First, there is the expression’s *reference*, i.e., its denotative relation to a *referent* (or denoted entity). Second, there is the expression’s *sense*, which Frege characterized as a particular manner in which the expression’s referent is cognitively presented to the language user. Frege motivated this distinction by drawing attention to sentences in which the meaning of a singular term is apparently *not* exhausted by its correlation with a referent. For example, if the meaning of a singular term were to consist only in what it refers to, then the true, but non-trivial, identity statement ‘The evening star is the morning star’ could not differ in meaning from the *trivially* true identity statement ‘The evening star is the evening star.’ Since the latter statement results from the former by substituting co-referential singular terms, any view that equates meaning with reference will necessarily fail to register any difference in meaning between the two statements. But the two sentences clearly *do* differ in meaning, since ‘The evening star is the morning star’ is not a trivial identity, but an informative identity – indeed, one that expresses the content of a genuine astronomical discovery – whereas ‘The evening star is the evening star’ is plainly trivial. Frege accounts for this by suggesting that while ‘the evening star’ and ‘the morning star’ have a common referent, they express different senses.

A language user therefore grasps the common referent differently in connection with each of the two expressions, and this in turn accounts for the difference in ‘cognitive value’ between the two identity statements. Frege applies the notion of sense to similar effect in addressing puzzles concerning the meaning of singular terms in so-called ‘intensional contexts,’ for example, belief reports.

In subsequent writings, Frege extended the sense-reference distinction beyond the category of singular terms (which, on Frege’s account, refer specifically to ‘objects’), to *all* categories of linguistic expression, including monadic and polyadic predicates (which refer to ‘concepts’ and ‘relations,’ respectively), and complete sentences. In the case of sentences, Frege identified as referents the two truth-values, ‘the true’ and ‘the false,’ and he characterized these as special ‘logical objects.’ A sentence’s sense is, by contrast, the ‘thought’ it expresses, where the thought is understood as a compositional product of the senses of the sentence’s linguistic subcomponents. As strained as this extension of the theory may appear, particularly with respect to reference, it brings to light two important features of Frege’s approach to meaning. First, it reflects his insistence that the theory of reference should comprise, *inter alia*, a theory of *semantic value* – that is, a systematic account of how the semantic values of complex expressions (which, in the case of sentences, are truth-values) are determined on the basis of the semantic values of their subordinate constituents. Second, it reflects an endeavor to integrate the theory of semantic value with a plausible general account of *linguistic understanding* (as given by the theory of sense). Seen in light of these general ambitions, Frege’s theory of sense and reference proposed an agenda that any comprehensive approach to the theory of meaning must in one way or another respect – a point that is amply borne out by subsequent developments in analytic philosophy of language.

Russell: Definite Descriptions and Logical Atomism

The idea that logical analysis forms the basis of a general philosophical method is central to the philosophy of Bertrand Russell (1872–1970). It is especially prominent in the works Russell produced over the first quarter of the twentieth century. In this period, Russell developed and defended the doctrine of ‘logical atomism,’ which grew out of his attempt to establish a version of logicism in the philosophy of mathematics, and came to encompass a wide variety of semantic, metaphysical, and epistemological ambitions. The common thread in Russell’s approach

to these matters consists in his emphasis on logical analysis as a method for clarifying the ontological structure of the world and the epistemological basis of our knowledge of it. As Russell put it, “the atom I wish to arrive at is the atom of logical analysis, not physical analysis” (1918: 37). Bound up with the notion of a logical atom, understood as a basic residue of logical analysis, is the notion of logical structure itself. Our aim in this section is to illuminate Russell’s conception of logical structure, or ‘logical form,’ as it emerges in his theory of linguistic meaning and in his broader atomism.

Russell’s theory of ‘definite descriptions,’ first articulated in his classic paper ‘On denoting’ (1905), paradigmatically illustrates Russell’s methodological reliance on logical analysis in addressing questions about meaning. The argument of the paper involves, among other things, the defense of a principle that Russell regarded as fundamental to the account of linguistic understanding. In *The problems of philosophy*, Russell gave the following succinct statement of the principle: “Every proposition which we can understand must be composed wholly of constituents with which we are acquainted” (1910: 32). At the time of ‘On denoting,’ Russell meant by a ‘proposition,’ roughly, the state of affairs that is expressed by an indicative sentence, whether or not that state of affairs actually obtains. A proposition’s ‘constituents’ are the real-world entities that figure in the state of affairs (or *would* figure in it, were the state of affairs to obtain). So understood, a proposition is not a linguistic entity, even in the attenuated sense of a Fregean thought. A proposition is, rather, a structured entity that comprises various nonlinguistic components of the world. What characterizes Russell’s principle of acquaintance as a principle of linguistic understanding, then, is not the linguistic *nature* of propositions, but the correlativity of propositions with the indicative sentences of a language. For Russell, understanding any such sentence requires direct experiential acquaintance with the non-linguistic constituents comprised in the proposition it expresses.

In ‘On denoting’ Russell addressed problems that the principle of acquaintance ostensibly confronts in connection with ‘denoting phrases’ – i.e., phrases of the form ‘some *x*,’ ‘every *x*,’ and especially ‘the *x*’ (i.e., so-called ‘definite descriptions’). Consider the statement: ‘The author of “On denoting” was a pacifist.’ Since Russell’s principle requires acquaintance with a proposition’s constituents as a condition for linguistic understanding, it would seem that only those personally acquainted with the author of ‘On denoting’ (i.e., with Russell himself) are in a position to understand the sentence. However, this highly counterintuitive consequence only arises on the assumption that the

denoting phrase ‘the author of “On denoting”’ functions as a genuine singular term, one that singles out Russell as a *constituent* of the corresponding proposition. Russell’s account of definite descriptions challenged this assumption by arguing that the characterization of definite descriptions as singular terms arises from a mistaken account of the logical form sentences containing them. According to this mistaken analysis, the sentence ‘The author of “On denoting” was a pacifist’ is an instance of the simple subject-predicate form *Ps*, where *s* indicates the occurrence of a singular term and *P* the occurrence of a predicate. Russell maintained that sentences containing definite descriptions have a far richer logical structure than this account would suggest. On Russell’s analysis, the statement ‘The author of “On denoting” was a pacifist’ is not a simple subject-predicate statement but has the form, rather, of a multiply quantified statement:

$$\exists x((x \text{ authored 'On denoting' } \& \forall y (y \text{ authored 'On denoting' } \rightarrow y = x)) \& x \text{ was a pacifist})$$

On this analysis, the statement says: there is an *x* such that (1) *x* authored ‘On denoting,’ (2) for any *y*, if *y* authored ‘On denoting,’ then *y* = *x* (this clause serving to ensure the uniqueness implied by the use of the definite article) and (3) *x* was a pacifist. So construed, the only nonlogical components of the sentence are the descriptive predicates ‘() authored “On denoting”’ and ‘() was a pacifist,’ with no trace remaining of the putative singular term ‘the author of “On denoting”’. Therefore, beyond an implicit understanding of the mechanisms of quantification and the logical relation of identity, acquaintance with the referents of these descriptive predicates suffices for understanding the sentence. The sentence still manages to be *about* Russell since he, and he alone, satisfies the descriptive predicates (or ‘propositional functions,’ in Russell’s terminology) contained in the sentence. However, it no longer singles out Russell as a *constituent* of the corresponding proposition, thus dispensing with the worry that the principle of acquaintance would require personal acquaintance with Russell as a condition for understanding what the sentence means.

The theory of definite descriptions vividly conveys the sense in which, for Russell, the surface grammar of natural language is inadequate as a guide to the analysis of logical form. Indeed, Russell maintained that many of the metaphysical and epistemological perplexities of traditional philosophy were a direct result of conflating the grammatical forms of natural language sentences with logical forms of the propositions we manage to express in natural language. In this connection, it is important to recognize that, for Russell, logical form is not a purely linguistic notion.

We have already taken note of the fact that Russell's early philosophy treats propositions as structured complexes that array various non-linguistic components of reality. Though Russell ultimately abandoned his early theory of propositions, he never abandoned the view that reality itself exhibits varieties of structure to which the details of a suitably perspicuous logical language must answer. In his lectures on *The philosophy of logical atomism* (1918), this view takes the form of a doctrine of 'facts,' where facts are understood as real-world complexes of individual objects and the properties and relations predicable of them. On Russell's characterization, a fact is a kind of complex that is inherently apt to determine a corresponding indicative statement as true or false – that is, true when the statement affirms the fact, and false when it denies it. The kernel of Russell's atomism consists in the view that the content of any statement (of whatever order of complexity) is ultimately analyzable in terms of the constellation of logically primitive facts that determine the statement as true or false. Russell's inventory of such facts includes 'atomic facts,' in which properties and relations are predicated of metaphysically 'simple' entities, and 'general facts,' which are facts concerning *all* or *some* of a particular category of entity. Atomic facts correspond to the atomic sentences, and general facts to the quantified sentences, of a logically regimented language. All other sentences are 'molecular' in the sense that they are compounds built up from atomic sentences and quantified sentences by the application of logical connectives such as 'not,' 'and,' 'or,' and 'if ... then ...'. Though molecular sentences assert neither atomic nor general facts, their truth or falsity is nevertheless *dependent upon* such facts in the sense that a molecular sentence will be determined as true or false *as a function of* the truth or falsity of its non-molecular subsentences. For example, if '*p*' and '*q*' are atomic sentences, then the disjunction '*p* or *q*' will be true just in case one or both of '*p*' and '*q*' are true, where the truth or falsity of these subsentences is determined directly by the atomic facts to which they relate.

Russell's metaphysical view of logical form – that is, his view that logical structure is an inherent characteristic of the facts that the real world ultimately comprises – is nicely expressed in a comment from his *Introduction to mathematical philosophy*. There Russell maintains that "logic is concerned with the real world just as truly as zoology, though with its more abstract and general features" (1919: 169). At least part of Russell's motivation for this 'substantive' conception of logic consists in his abiding conviction that the structure of language (or of an ideal language, at any rate) and the structure of the world must in

some way coincide if there is to be any prospect of expressing our knowledge of the world by linguistic means. The task of giving a detailed account of this community of form between language and world was one that Russell wrestled with many times over, but which he ultimately left to the talents of his most gifted student, Ludwig Wittgenstein – the second great exponent of the philosophy of logical atomism.

Wittgenstein on Logic and Language

Of the figures we are discussing, arguably Wittgenstein (1889–1951) addressed the question of the relation between logic and language most extensively. His earliest major work, the *Tractatus logic-philosophicus*, devotes much attention to this problem. It is on this early work that we will focus here. In this work he praised Russell for discovering that the apparent logical form of a proposition need not be its real logical form. He supplemented Russell's view with the claim that the real form of a proposition is a picture of a state of affairs in the world.

Propositions, according to the *Tractatus*, are pictures of facts. The structure of a proposition mirrors the structure of the fact it represents. What a fact and the proposition that describes it have in common is their form. "The picture, however, cannot represent its form of representation; it shows it forth" (2.172). Here we see the important Tractarian distinction between *saying* and *showing*. The statement 'it is now raining' *says* something about the world. The statement 'it is either the case that it is now raining or it is not the case that it is now raining' *says* nothing about the world. It does, however, *show* the logical relations between facts. If something can be shown it cannot be said (4.1212). It follows that nothing concerning the logical relations between facts can be said.

According to the *Tractatus* "the world is the totality of facts, not of things" (1.1). That is to say, the world is not completely described by a list of all the objects that it contains. Rather, a complete description of the world would consist of all true sentences. Facts can either be atomic or compound, and correspondingly there are two types of propositions. Atomic facts are the most basic type of fact, and all atomic facts are independent of one another. Likewise, any possible set of atomic propositions could be true at the same time. This does not hold generally, as *p* and $\sim p$ (it is not the case that *p*) cannot both be true at the same time. Compound propositions are built up by *truth functions* on atomic proposition. Any operator, including the logical operators (and, or, not, ...), that takes sentences as arguments and assigns a truth value to the compound expression based only on the truth value of the arguments, is called a truth

functional operator. For instance, 'or' designates a truth function with two argument places: the truth value of the sentence 'p or q' depends only on the truth value of the sentences 'p' and 'q'. On the other hand, in the sentence 'Julius Caesar conquered Gaul before Rome fell to barbarians,' '... before ____' designates a function that takes sentences as arguments, but it is not truth functional since we need to know more than the truth value of the arguments to determine the truth value of the compound. Wittgenstein observed that all propositions are either atomic or built up by truth functions on atomic propositions. Because of this all propositions can be expressed as a truth function on a set of atomic propositions.

Statements such as all of those of the form 'p or \sim p' are tautologies: they are true no matter what the truth value of the constituents. We can know for certain that a tautology is true, but this is only because tautologies are true independently of which atomic facts turn out to be true (and because all sentences are truth functions of atomic sentences). We cannot say that the world has a certain logical structure, this can only be shown. It is tautologies that show the logical syntax of language, but tautologies say nothing. "Logical propositions describe the scaffolding of the world, or rather they present it. They 'treat' of nothing" (6.124).

Concerning sentences of natural language, Wittgenstein thought that no serious reconstruction is necessary. "In fact, all the propositions of our everyday language, just as they stand, are in perfect logical order" (5.5563). Furthermore, Wittgenstein thought that what he says about the logical structure of language must already be known by anyone who can understand the language. "If we know on purely logical grounds that there must be elementary propositions, then everyone who understands propositions in their unanalysed form must know it" (5.5562). Remember that Wittgenstein, following Russell, distinguished between the apparent logical form of a proposition and its real logical form. The logical form of natural languages is extremely complex and shrouded in conventions. "Man possesses the capacity of constructing languages, in which every sense can be expressed, without having an idea of how and what each word mean – just as one speaks without knowing how the single sounds are produced. Colloquial language is part of the human organism and is no less complicated than it. From it it is humanly impossible to gather immediately the logic of language" (4.002).

While ordinary use of natural language is in perfect logical order, philosophy arises from the abuse of natural language. Wittgenstein thinks that philosophy is nonsense because it attempts to state what cannot be said. "Most propositions and questions,

that have been written about philosophical matters, are not false, but senseless" (4.003). The view that philosophy as standardly practiced is senseless, and therefore that a radically new approach to philosophy must be developed had a profound influence on a group of philosophers who held weekly meetings in Vienna – the Vienna Circle.

Carnap and the Vienna Circle

Rudolf Carnap (1891–1970) is generally regarded as the most influential member of the Vienna circle. This group (often called the logical positivists or logical empiricists) studied Wittgenstein's *Tractatus* carefully and much of what they wrote was inspired by or was a reaction to this work. Logical positivism is often thought of as being characterized by its commitment to *verificationism*. In its strictest form, verificationism is the view that the meaning of a sentence consists in the method of its verification – that is, in the epistemic conditions under which the statement would properly be acknowledged as true. In a less strict form, it is the view that the meaning of a sentence consists of what would count as evidence for or against it. There was much debate in the circle as to what form the verificationist principle should take. There was in the circle very little objection (Gödel being the notable exception) to the thesis that there are two different kinds of statements – empirical (synthetic) and logico-mathematical (analytic) statements. Concerning empirical statements, their meaning is given by what would amount to a verification (or confirmation on a less strict view) of the statement or its negation. Concerning logico-mathematical statements, the circle was much influenced by Wittgenstein's view that tautologies are *a priori* truths – truths that are knowable independently of experience because they say nothing concerning the state of the empirical world.

What Wittgenstein counted as a logical truth (a tautology) was not sufficiently broad to include all of mathematics. Since mathematical truths are not empirical assertions, members of the Vienna circle thought they should have the same status as other logical truths. Carnap undertook to broaden the definition of logical truth so as to include all mathematical statements. To do this Carnap had to answer the question of what makes something a logical truth. Carnap's answer to this question involved the adoption of a strong form of conventionalism, which he expressed in terms of his famous 'principle of tolerance': "In logic, there are no morals. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes. All that is asked of him is that, if he wishes to discuss it, he must state his methods

clearly, and give syntactic rules instead of philosophical arguments" (*The logical syntax of language*, §17). This principle states that logical truth is a matter of convention. Which statements are treated as belonging to the set of analytic statements is a matter of pragmatic decision, provided that the set can be clearly defined. There is, for Carnap, no logical structure of the world that is either rightly or wrongly captured by our choice of logic. Logical relationships between sentences are a matter of stipulation on our part. However, by classifying logical statements as analytic, and therefore independent of empirical circumstances, Carnap preserves the Wittgensteinian idea that logical truths say nothing about the world. Carnap later summarized his position on logico-mathematical truth by claiming that analytic statements are true in virtue of their meaning.

Carnap's principle of tolerance was inspired by the debates concerning the foundations of mathematics in the 1920s. One party in this debate were the intuitionists, who did not believe that we have grounds to assert a mathematical sentence of the form 'p or \sim p,' unless we have a proof of either p or of \sim p. According to classical logic, the sentence 'p or \sim p' is a tautology; it therefore stands in no need of prior justification. Intuitionists, therefore needed to abandon classical logic in favor of a logic that would not count all instances of the law of the excluded middle (p or \sim p) as valid. Carnap saw both classical and intuitionistic logic as well motivated, and saw nothing that could decide between the two. He therefore saw the decision of which logic to adopt as a matter of choice.

Further developments in logic amplified the differences between Carnap and Wittgenstein. Gödel's famous incompleteness theorems made use of a technique that has since become known as Gödel numbering. By Gödel numbering we assign code numbers to expressions of the language. Through this coding technique, a language capable of expressing arithmetical properties becomes a device for discussing certain syntactic properties of any language system. Carnap saw this as a refutation of Wittgenstein's idea that the logical syntax of language is inexpressible. In fact, one of the general goals of Carnap's *The logical syntax of language* was to show that it is possible to deal in a clear systematic manner with the syntactic properties of any language. Recall that for Wittgenstein, we cannot say anything concerning the logical syntax of language.

Carnap's logical tolerance led him to assert that even statements of the form $(\exists x)Px$, which assert the existence of an object with the property P, might be true by stipulation. That we could stipulate an object into existence seemed odd to many philosophers.

In order to address this worry, Carnap formulated a distinction between *internal* and *external* questions of existence. In the language system of arithmetic it is provable that $(\exists x) (7 < x < 9)$. Relative to this language, the question of the existence of numbers is trivial. But when someone asks whether numbers exist they do not mean to be asking the questions in such a manner that it is answerable in by appeal to the standards of proof and disproof that prevail in arithmetic. Rather, they mean to ask if the numbers really exist in some absolute sense. Carnap viewed such 'external' questions as unanswerable given that they remove the questions from a context in which there are clear standards for addressing it, without embedding it in another context where there are any such standards. But the coherence of the language system that includes, for instance, the numbers does not depend on a positive answer to the external question of the existence of numbers. In this way, not only the logical structure of the world but its ontology as well becomes a matter of convention.

Quine: the Thesis of Gradualism

W. V. O. Quine (1908–2000) began his philosophical career as a self-described disciple of Carnap's. However, from his earliest interaction with Carnap, Quine questioned Carnap's strict division between analytic and synthetic sentences. This early reaction to Carnap's work grew into a major break between the two philosophers. Recall that the analytic/synthetic distinction divides sentences (respectively) into those that concern the world and are capable of being empirically confirmed, and those that are accepted by stipulation and are true in virtue of meaning. Quine thought that this difference in kind ought to be replaced with a difference of degree. This is known as the thesis of gradualism.

For Quine our knowledge forms a structure like a web. The nodes of the web are sentences and the links between nodes are entailment relations. Only at the periphery are our decisions to accept or reject sentences directly influenced by experience. Decisions over sentences closer to the center are of an increasingly 'theoretical' character, with accepted logical and mathematical statements forming the most central class. The ordering of sentences from periphery to interior is based on how willing we would be to abandon a sentence when revising our beliefs in light of new evidence. For sentences like 'this table is red' we can easily imagine a set of experiences that would lead us to abandon it. By contrast, it is far more difficult to imagine the experiences that would lead us to abandon ' $2 + 2 = 4$.' Abandoning this statement would entail a far more radical change in our

overall belief system. However, for Quine, the difference is one of degree, rather than kind. No sentence, mathematical, logical or otherwise, is ultimately immune from revision in light of experience. Physics, for instance, in departing from Euclidean geometry has abandoned sentences such as 'between any two points exactly one straight line can be drawn' once believed to be on the firmest foundation.

We have seen that, for Carnap, logico-mathematical truths are not responsible to any aspect of the world. We are perfectly free to accept any set of sentences to count as analytic. The way the world is affects only the practical utility of our choices of analytic statements; it does not affect the theoretical legitimacy of those choices. (However, Carnap is far more interested in giving reconstructions of existing notions instead of constructing arbitrary systems.) For Quine, on the other hand, logical and mathematical truths are on par with highly theoretical statements of physics. It may turn out that by abandoning classical logic or altering our mathematics we will be able to formulate more simple scientific theories. Since simplicity is one of the norms of theory choice, it may be that our best scientific theory does not conform to the laws of classical logic.

Carnap's principle of tolerance suggests that logical truths are true by virtue of the meanings assigned to the logical vocabulary. Quine rejects this view and sees logical truths as subject to the same standards of acceptance as any other scientific claim. Since there are a certain sets of experiences that would lead us to reject what we now regard as a logical truth, Quine maintained that we could no longer hold, as Wittgenstein did, that logical truths are true independently of how things happen to be in the empirical world. Logical truths therefore lose their special status and become statements on par with other scientific claims. They are true because they are part of our best description of the world.

See also: A Priori Knowledge: Linguistic Aspects; Analytic Philosophy; Analytic/Synthetic, Necessary/Contingent, and a Priori/a Posteriori: Distinction; Logical Consequence; Propositions; Semantic Value; Sense and Reference: Philosophical Aspects.

Bibliography

- Boole G (1854). *The laws of thought*. [Reprinted New York: Dover, 1958.]
- Carnap R (1937). *The logical syntax of language*. Amethe Smeaton (trans.). New Jersey: Littlefield Adams, 1959.
- Carnap R (1950). 'Empiricism, semantics and ontology.' In Sarkar S (ed.). 1996.
- Frege G (1879). 'Begriffsschrift: a formula language modeled upon that of arithmetic, for pure thought.' Bauer-Mengelberg S (trans.). In van Heijenoort Jean (ed.) *From Frege to Gödel: a source book in mathematical logic*. Cambridge, MA: Harvard University Press, 1976.
- Frege G (1892). 'On sense and reference.' Black M (trans.). In Black M & Geach P T (eds.) *Translations from the philosophical writings of Gottlob Frege*, 2nd edn. Oxford: Basil Blackwell, 1960.
- Frege G (1893). *Grundgesetze der Arithmetik* (vol. 1). Jena: Verlag Hermann Pohle. [Partially translated as Furth M, *The basic laws of arithmetic* by University of California Press, Berkeley 1964.]
- Leibniz G W (1696). 'Letter to Gabriel Wagner on the value of logic.' Loemker D L (ed. & trans.). In *Philosophical papers and letters*, 2nd edn. Dordrecht & Boston: Reidel Publishing, 1976.
- Leibniz G W (1679). 'Elements of a calculus.' In Parkinson G H R (ed. & trans.). 1966.
- Leibniz G W (1679/1686). 'A specimen of the universal calculus.' In Parkinson G H R (ed. & trans.). 1966.
- Parkinson G H R (ed. & trans.) (1966). *Logical papers*. Oxford: Clarendon Press.
- Quine W V (1951). 'Two dogmas of empiricism.' In Sarkar S (ed.). 1996.
- Quine W V (1969). 'Epistemology naturalized.' In *Ontological relativity and other essays*. New York: Columbia University Press.
- Russell B (1905). 'On denoting.' [Reprinted in Marsh R C (ed.) *Logic and knowledge*. London: Unwin, 1988.]
- Russell B (1912). *The problems of philosophy*. [Reprinted London: Oxford University Press, 1986.]
- Russell B (1918). *The philosophy of logical atomism*. Pears D (ed.). La Salle, Illinois: Open Court, 1972.
- Russell B (1919). *Introduction to mathematical philosophy*. New York: Clarion, 1971.
- Sarkar S (ed.) (1996). *Science and philosophy in the twentieth century v. 5*. New York: Garland.
- Wittgenstein L (1918). *Tractatus logico-philosophicus*. Ogden C K (trans.). London: RKP, 1988.

Logical Consequence

P Blanchette, University of Notre Dame, Notre Dame, IN, USA

© 2006 Elsevier Ltd. All rights reserved.

Fundamentals

Logical consequence is the relation that holds between the premises and conclusion of an argument when the conclusion *follows from* the premises, and does so for purely logical reasons. When a conclusion is a logical consequence of premises, the truth of those premises suffices to guarantee the truth of the conclusion. To clarify, we'll look at some examples.

When we reason that

(A1) Socrates is mortal

follows from

(A2) Socrates is human

and

(A3) All humans are mortal,

we need not appeal to any known facts about Socrates, about humanity, or about mortality. These specifics are irrelevant to the fact that (A1) follows from (A2) and (A3), which shows that the sense of 'following-from' involved here is the purely logical sense. That is, (A1) is a *logical* consequence of (A2) and (A3).

By contrast, when we reason that

(B1) There are mammals in the ocean

follows from

(B2) There are dolphins in the ocean

we must appeal to facts peculiar to the nature of mammals and of dolphins. We appeal, specifically, to the fact that dolphins are mammals. In this case, although there is a sense in which (B1) 'follows from' (B2), (B1) does not follow *logically* from (B2). It follows, one might say, *biologically*, because an appeal to biological facts is needed to get from (B2) to (B1). Nevertheless, the fact that (B1) follows in this extra-logical way from (B2) is because of the relation of logical consequence. Specifically, it is because of the fact that (B1) is a logical consequence of (B2) together with

(B3) All dolphins are mammals.

That (B1) is a *logical* consequence of (B2) and (B3) can be seen by noting that it follows from them independently of the specific nature of the objects, properties, and relations mentioned in these statements.

In general, all cases of 'following from' are due, in this way, to the relation of logical consequence. If a conclusion follows from some collection of premises, this is because that conclusion is a *logical* consequence of the premises, together perhaps with various ancillary claims that are presupposed in the given context.

Logical consequence is therefore a ubiquitous relation: all of our reasoning turns on recognizing (or attempting to recognize) relations of logical consequence, and virtually all of the important connections between theories, claims, predictions, and so on, are in large part due to logical consequence. Furthermore, whenever we say that a given argument is *valid* or that it is *invalid*, or that a particular set of claims is *consistent* or *inconsistent*, we are employing the notion of logical consequence: a valid argument is one the conclusion of which is a logical consequence of its premises, whereas a consistent set of claims is a collection that has no contradiction as a logical consequence. Because the central logical notions of validity, consistency, etc., are definable in terms of logical consequence, the investigation of the nature of logical consequence is at the same time the investigation of the nature of the logical properties and relations in general.

The Formal Study of Logical Consequence

The modern investigation of logical consequence is closely connected to the discipline of *formal logic*. Formal logic is the study of formal (i.e., syntactically specified) languages, and of various philosophically and mathematically significant properties and relations definable in terms of such languages. Of particular significance for the study of logical consequence are two kinds of relations definable on the formulas of a formal language, the relations of *proof-theoretic consequence* and of *model-theoretic consequence*.

Given a formal language, a relation of proof-theoretic consequence is defined via the rigid specification of those sequences of formulas that are to count as *proofs*. Typically, the specification is given by designating specific formulas as *axioms*, and designating some *rules of inference* by means of which formulas are provable one from another. Both axioms and rules of inference are specified entirely syntactically. A proof is then a series of formulas each of which is either taken as premise, or is an axiom, or is obtained by previous formulas in the series via a rule of inference. A formula φ is a proof-theoretic consequence of a set Σ of formulas if and only if there is a proof the premises of which are

among the members of Σ , and the conclusion of which is φ .

Model-theoretic consequence, by contrast, is defined in terms of a range of *interpretations* (or *models*) of the formal language in question. While the vocabulary of the language is divided into the 'logical' terms (typically, analogues of the English-language 'and,' 'or,' 'not,' 'if...then,' and 'for all'), the meaning of which is taken as unchanging, and the 'non-logical' terms (typically analogues of natural-language predicates and singular terms), an *interpretation* is an assignment of objects and sets of objects to the non-logical terms. In the standard case, the formulas are taken to have a truth-value (i.e., to be either true or false) on each such interpretation. A formula φ is then a model-theoretic consequence of a set Σ of formulas if and only if there is no interpretation on which each member of Σ is true while φ is false.

The connection between these defined relations and logical consequence arises when the formulas in question are taken to stand as representatives of natural-language sentences or the claims they express. Given such a representation-relationship, the relations of proof-theoretic and of model-theoretic consequence are typically designed so as to mirror, to some extent, the relation of logical consequence. Thus, the idea behind a standard design of a relation of proof-theoretic consequence is that it only count as axioms those formulas representing 'logical truths' (e.g., 'Either 5 is even or 5 is not even'), and that its rules of inference similarly mirror logical principles. In such a case, a formula φ will be a proof-theoretic consequence of a set Σ of formulas only if the kind of ordinary sentence represented by φ is indeed a logical consequence of the ordinary sentences represented by the members of Σ . This does not ensure that the relation of proof-theoretic consequence exhausts the relation of logical consequence, for two reasons: first of all, the formal language in question may not contain representatives of all ordinary sentences; second, the proof system may not be rich enough to reflect all of the instances of logical consequence amongst even those ordinary sentences that are represented in the language. The system of proof-theoretic consequence will, however, have the virtue of being well defined and tractable. Similar remarks apply to the relation of model-theoretic consequence: in a well-designed formal language, the relation of model-theoretic consequence will mirror, in important ways, the relation of logical consequence. The intention in designing such a system is, typically, that φ will be a model-theoretic consequence of Σ if and only if the kind of ordinary sentence represented by φ is a logical consequence of those represented by the members of Σ .

Given a particular language together with its proof-theoretic and model-theoretic consequence relations, the question arises whether those relations are co-extensive: whether, that is, φ is a proof-theoretic consequence of Σ if and only if φ is a model-theoretic consequence of Σ . In some cases, the answer is 'yes,' and, in some, 'no.' Each half of the inclusion is a separate, significant issue: when every proof-theoretic consequence of each set of formulas is also a model-theoretic consequence of that set, the system is said to be *sound*, and when every model-theoretic consequence of each set of formulas is also a proof-theoretic consequence of that set, the system is said to be *complete*. The soundness of a system is typically a straightforward matter, following immediately from the design of the proof-theoretic system; completeness is typically a considerably more significant issue. The most-important system of logic, that of classical first-order logic, was proven by Kurt Gödel in 1930 to be complete; this is the celebrated 'completeness theorem for first-order logic.' First-order logic is, in various ways, the 'strongest' complete system (see Enderton, 1972).

Formal systems, i.e., formal languages together with proof-theoretic or model-theoretic consequence relations, differ from each other in a number of ways. Most important for the purposes of the study of logical consequence are the following two differences: (1) proof-theoretic relations differ over the axioms and rules of inference they include, and hence over the instances of logical consequence that they represent. Some such differences are just because the languages of some such systems are expressively weaker than others, so that principles contained in one simply cannot be expressed in the other. More interesting are differences motivated by differing views of logical consequence itself. Thus, for example 'classical' logic differs from intuitionist logic in including the principle of *excluded middle*, the principle guaranteeing the truth of all statements of the form *p-or-not-p*. As the proponent of intuitionist logic sees it, this principle is not universally accurate, and hence should not be included in a system of logic. (2) Model-theoretic relations differ in a number of small ways, including the specifics of the definition of *interpretation*, and of the definition of truth-on-an-interpretation. More important, the model-theoretic consequence relations for different systems differ when the formal languages in question are importantly structurally different. Thus, for example, standard *second-order logic* has a richer model-theoretic consequence relation than does first-order logic, and there are natural-language arguments whose second-order representation yields a conclusion that is a

model-theoretic consequence of its premises, but whose first-order representation does not (see van Dalen, 2001; Shapiro, 1991).

The question of the extent to which each such system gives an accurate characterization of logical consequence is of central philosophical concern. With respect to the relations of proof-theoretic consequence, debate turns on the accuracy of specific axioms and rules of inference. With respect to relations of model-theoretic consequence, the significant debate is rather over the question of the extent to which model-theoretic consequence relations in general (or, perhaps, that relation as applied to classical first-order logic) offer an *analysis* of the ordinary, non-formal relation of logical consequence. If logical consequence is in some sense ‘essentially’ the relation of truth-preservation across interpretations, then model-theoretic consequence has a privileged position as simply a tidied-up version of the core relation of logical consequence. If, by contrast, the relation of truth-preservation across interpretations is simply another sometimes-accurate, sometimes-inaccurate means of representing the extension of the relation of logical consequence, then model-theoretic consequence has no immediate claim to accuracy (see Etchemendy, 1990).

General Philosophical Concerns

In addition to questions surrounding its appropriate formal representation, the investigation of logical consequence includes questions concerning the nature of the relation itself.

One important cluster of such questions concerns the relata of the relation. Here we want to know whether the items between which logical consequence holds are, say, the *sentences* of ordinary language, or the non-linguistic *propositions* expressed by such sentences, or something else altogether. Although logical consequence is perhaps most straightforwardly viewed as a relation between sentences, one reason to reject this idea is that sentences, at least when thought of as syntactic entities (strings of letters and spaces), seem the wrong kinds of things to bear that relation to one another. Because any given sentence so understood could, under different circumstances, have had a quite different meaning, and would thereby have borne different logical relationships to other sentences, it is arguable that the sentence itself is not the primary bearer of this relation but is, rather, just a means of expression of the primary bearer. This line of reasoning motivates the view of non-linguistic propositions, the kinds of things expressed by (utterances of) fully interpreted sentences, as the relata of

logical consequence. The central reason for rejecting *this* proposal, though, is skepticism about the existence of such things as nonlinguistic propositions. A third option is to take the relata of the logical consequence relation to be sentences-in-use, essentially pairs of sentences and meaning-conferring practices (see Cartwright, 1987; Strawson, 1957; Quine, 1970).

The second, related collection of questions concerning logical consequence arises from the inquiry into what *makes* one thing a logical consequence of others. Here, we are looking for an explanation or an analysis of logical consequence in terms of other, more well-understood notions. One potential answer is that logical consequence is to be explained in terms of the meanings of various specific parts of our vocabulary, specifically in terms of the meanings of the ‘logical’ words and phrases (see above). A second, not necessarily competing, account is that logical consequence is because of the *form*, or overall grammatical structure, of the sentences and arguments in question. A third type of answer, mentioned above, is that logical consequence is best explained in terms of model-theoretic consequence. Various of the accounts of logical consequence have been criticized on grounds of circularity: to say that φ ’s being a logical consequence of Σ is *because of* some other relation between φ and Σ is, arguably, to say that the claim that φ is a logical consequence of Σ is itself a *logical consequence* of the purported explanans. If this charge of circularity is accurate, it is arguable that all such explanations of the nature of logical consequence will be found to be circular, with the result that this relation must be taken to be ‘primitive,’ not capable of reduction to anything else. Part of the debate here will turn on what one takes the nature of explanation to be, and on whether explanation requires reduction (see Quine, 1936).

In short: although it generally is agreed that some claims are logical consequences of others, there is scope for important disagreement about (a) which specific claims are in fact logical consequences of which others, (b) how to construe the notion of ‘claim’ involved here, and (c) how to give a correct account of the nature of the relation of logical consequence. Because of the connections between these issues and general positions in the philosophy of logic, philosophy of mathematics, and philosophy of language, one’s preferred answers to the questions noted here will turn in large part on one’s position with respect to a host of surrounding topics.

See also: Logic and Language: Philosophical Aspects; Logical Form in Linguistics; Propositions.

Bibliography

- Blanchette P A (2000). 'Models and modality.' *Synthese* 124(1), 45–72.
- Blanchette P A (2001). 'Logical consequence.' In Goble L (ed.) *The Blackwell guide to philosophical logic*. Malden, MA/Oxford: Blackwell Publishers. 115–135.
- Cartwright R (1987). 'Propositions.' In Butler R J (ed.) *Analytical philosophy, 1st series*. Oxford: Blackwell. Reprinted in Cartwright R. *Philosophical essays*. Cambridge and London: MIT Press, 1987. 33–53.
- Enderton H (1972). *A mathematical introduction to logic*. Orlando, FL: Academic Press.
- Etchemendy J (1990). *The concept of logical consequence*. Cambridge, MA: Harvard University Press. Reprinted 1999, Stanford: CSLI Publications.
- Goble L (ed.) (2001). *The Blackwell guide to philosophical logic*. Malden, MA/Oxford: Blackwell Publishers.
- Quine W V O (1936). 'Truth by convention.' In Lee O H (ed.) *Philosophical essays for A. N. Whitehead*. New York: Longmans. Reprinted in Quine W V O. *The ways of paradox and other essays*. Cambridge, MA/London: Harvard University Press, 1976. 77–106.
- Quine W V O (1970). *Philosophy of logic*. Englewood, NJ: Prentice Hall.
- Shapiro S (1991). *Foundations without foundationalism: a case for second-order logic*. Oxford: Oxford University Press.
- Strawson P F (1957). 'Propositions, concepts, and logical truths,' *Philosophical Quarterly* 7. Reprinted in Strawson P F. *Logico-Linguistic Papers*. London: Methuen & Co. 1971. 116–129.
- Tarski A (1936). 'On the concept of logical consequence,' translation of 'O pojęciu wynikania logicznego.' In *Przegląd Filozoficzny*, 39, 58–68. English translation in *Logic, semantics, metamathematics* (2nd edn.). Woodger J H (trans.) & Corcoran J (ed.). Indianapolis: Hackett Publishing Company, 1983. 409–420.
- van Dalen D (2001). 'Intuitionistic logic.' In Goble L (ed.) *The Blackwell guide to philosophical logic*. Malden, MA/Oxford: Blackwell Publishers. 224–257.

Logical Form in Linguistics

D Blair, University of Western Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

To describe the logical form of some claim is to describe its logically significant properties and structure, showing its connection to other claims via what it entails and what it entails it. Given the variety of claims that philosophers have taken in an interest in, it is not surprising that there are a large number of theories of logical form. But even if there is no shortage of theories aiming at the logical form of, e.g., propositional attitude sentences or counterfactual conditionals, surprisingly little attention has been given to the prior question of **what** logical form is to begin with. Just as importantly, it is not clear what it is that is supposed to **have** a logical form in the first instance. Is it, for example, a linguistic object like a sentence, or the utterance of a sentence, or something different from both of these, such as the proposition expressed by an utterance of a sentence?

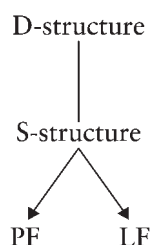
The presence of logic within the notion of logical form may make one suspicious of paying too much attention to the details of natural language. Other kinds of items seem better suited to having logical forms. For example, propositions have whatever truth conditions they have essentially, whereas sentences do not: 'snow is white' might have meant that most colorless beverages lack sodium. Further, it is a

notorious fact about natural language that it contains a good deal of vagueness and context sensitivity that is hard to capture within a theory of inference. Facts like these have made philosophers wary of placing too much emphasis on natural language sentences. At the very least, one would want to purge natural language of its logically problematic features before building upon it a theory of logical form. This was precisely the reaction of Frege (1952) and Russell (1919) to the defects of natural language. For them, one needed to formulate an ideal language free from the flaws of natural language in order to spell out the content of various claims. Only then could one think about constructing theories of logical form. Frege's *Begriffschrift* formulated an ideal language in which to conduct arithmetic and overcame some of the difficulties of explaining inferences involving multiple quantifiers that beset earlier logical theories.

But even if having a logically perspicuous representation of the propositional content of an assertion makes it easier to assess how well a theory accords with what is said about, e.g., the good or the propositional attitudes, there are serious questions concerning how such representations are related to the grammatical properties of a sentence. In the hands of Frege and Russell, one simply translated, as best one could, from natural language into an ideal language. These languages were specifically designed to expedite inference, and so no question arises about their logical forms.

But until the last few decades, the kinds of structures required for the purposes of detailing the inferential properties of natural language sentences were thought to be quite remote from anything one might call ‘the grammar’ of a language. Indeed, one way of **motivating** talk of logical form was by showing the deficiencies of theories of meaning built upon generalizations of apparent grammatical form and function.

A number of developments in the 1960s and 1970s changed this picture. A growing number of philosophers became intrigued with the idea of constructing theories of meaning for natural languages directly. The idea that such a theory could be done **systematically** stems in large part from the work of Noam Chomsky in the 1950s and 1960s, showing how rigorous theories of grammatical structure were possible. In light of the success of Chomsky’s program, it was natural to wonder whether a semantic theory along the lines of his work in syntax could be constructed. The classic picture of the grammatical structure of a sentence involves a series of levels of representation, the most well known of which is the so-called ‘T-model.’ In this model, there are four ‘levels of representation’: D-structure, S-structure, LF, and then PF, or the phonological form of a sentence. Since the last item is a representation of a sentence’s phonological properties, I leave it aside. Each level is related to the one before via the application of a rule or set of rules. The conception of rules has changed over the years, but the underlying idea is that syntactic structure of a sentence is built up, step by step, through a series of representations, each having its own properties. Diagrammatically, what we have is the following:



The ‘S-structure’ or surface structure of a sentence is what corresponds, nearly enough, to the order of expressions as heard or written. ‘LF’ or logical form is a syntactic representation that is derived from the S-structure via a set of transformations, just as S-structures were derived from D-structures via transformations.

Since only one level of representation seems to correspond to the overt form of a sentence, it follows that a good deal of syntactic structure remains hidden. The idea that unpronounced structure can

be given a grammatical motivation is compelling. Consider the following pair of sentences:

- (1) John kissed Mary
- (2) Who did John kiss

The leftmost WH-phrase in (2) is intuitively related to the position of ‘Mary’ in (1). The grammar of English disguises this fact by requiring that unstressed WH-phrases in sentences like (2) be fronted. Were English different in this regard, the parallel would be more obvious. Interestingly, a good many languages allow for just this possibility while others require all WH-phrases to be placed at the left- periphery of a sentence. A more perspicuous representation of English would abstract from these kinds of provincial eccentricities of surface form and expose, via a logically perspicuous notation, just these parallels.

There is evidence that the grammatical structure of sentences like these in different languages is abstractly identical, i.e., that all WH-phrases are located at the edge of a clause at some level of representation. In some languages, like Russian, this is overtly true, even when there are several WH phrases in the clause. In other cases, like Chinese, there is little or no movement to the edge of the clausal periphery (see Huang, 1982). The difference between the overt forms of WH-questions then doesn’t disguise just the logical or semantic structure of a sentence; it hides the grammatical structure as well. A more articulated version of (2) shows this abstract structure:

- (3) Who did John kiss *t*

The key idea is that movement of a WH-phrase may occur overtly, as in English, or ‘covertly,’ as in some cases of French. When the WH-phrase does move, however, what we end up with is (3) The movement of the WH-phrase to its position at the left edge of the clause leaves a record in the form of a ‘trace,’ notated above as ‘t.’

Structures like (3) resemble, in a rather striking way, the kinds of representations that one finds within first-order logic, in particular with respect to the relationship between a quantificational expression and a variable that it binds. Let’s look at this in more detail. It is now commonplace to use examples of scope ambiguities as evidence for the ambiguity of sentences, one to be sorted out in a semantic theory. Thus, a sentence like (4) is ambiguous depending upon whether or not one takes the quantificational phrase ‘every boy’ to have scope over the subject quantificational phrase ‘some girl’ or vice versa, i.e., (5a/b):

- (4) Some girl danced with every boy

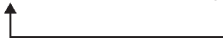
- (5a) $\exists x: \text{girl}(x) [\forall y: \text{boy}(y) [\text{danced}(x,y)]]$
 (5b) $\forall y: \text{boy}(y) [\exists x: \text{girl}(x) [\text{danced}(x,y)]]$

The usual way of describing this difference is to say that in (5a), ‘some girl’ has scope over ‘every boy,’ while in (5b), the opposite relation holds. The scope of the quantifiers is determined by looking at the material appearing to its right, i.e., the closest formula that does not contain the expression within the first order translation.

It turns out that one can define the relevant relation in syntactic terms as well, using the properties of phrase structure. To see this, consider the core syntactic relation of c-command.

An expression α c-commands an expression β if and only if the first branching node dominating α dominates β and neither α nor β dominates the other.

What is important is that one can use this definition to say something about quantificational scope. Suppose we take quantificational expressions to move to positions from which they c-command their original position:

[ZP XP [ZP ... [... [... t_i ...]...]


In this case, XP c-commands ZP and everything that is contained in the latter, including the trace of XP. Strikingly, when we look at what the structure of 1 is when this structure is explicit, we see the kind of structure required for the definition of scope:

- (6) [_S [_{QP} Some girl]₂ [_S [_{QP} Every boy]₁
 [_S t_2 [_{VP} danced t_1]]]]

For the reading of (4) where the scopes of the quantificational NPs are inverted relative to their surface order, ‘every boy’ is adjoined to a position from which it c-commands both ZP and the position to which ‘some girl’ has been adjoined:

- (7) [_S [_{QP} Every boy]₁ [_S [_{QP} Some girl]₂ [_S t_2 [_{VP}
 danced t_1]]]]

Both of these movements can be given more detailed defense; see May (1977). The structures that seem to be needed for semantics and that philosophers have thought were disguised by ordinary grammar really are hidden, although not quite in the way they thought. What is hidden is more syntactic structure.

Of course, ‘LF’ is a *syntactic* level of representation and is not a semantic representation. This is not to suggest, however, that no gain has been made within theorizing about natural language by incorporating the LF hypothesis. For one could hold that the grammatical structures that are interpreted by the semantic

theory are just those provided by a theory of grammar incorporating the LF hypothesis. There is no need to first regiment the formal structures of sentences into something to which semantic rules could then apply. What one finds in the idea of LF is the idea that natural languages already have enough structure to supply a lot of what is needed for the purposes of semantics.

Further developments within syntactic theory have made the concept of logical form more prominent. Thus, Chomsky (1995) and others have proposed that the **only** level of grammatical representation is LF, although the role of LF is likely to change, just as it has in the past (see, e.g., Lasnik, 2001). Even so, it is apparent that progress has been made in joining together two bodies of thinking about language, one rooted in traditional philosophical problems about the representation of logic and inference and the other in more recent developments coming from linguistics.

There are limits, however, to how much philosophical work a linguistic-based approach to logical form can do. Recall that one of the problems that has made many philosophers wary of paying too much attention to natural language concerned such things as the context sensitivity of certain aspects of natural language sentences. It is an open question just how to treat different kinds of context sensitivity within natural language, and whether revisions are needed to our conception of logical form in natural language in order to accommodate it. It is also true that a good number of philosophical projects targeting logical form are usually concerned with the conceptual analysis of certain notions, e.g., moral goodness, knowledge, etc. Indeed, one of the traditional roles of logical form within philosophy is to serve as scaffolding for just these sorts of projects. Doubts about the viability of conceptual analysis to one side, this is what has given weight to the claim that ‘ordinary language’ disguises the logically significant structure of our concepts. But if this is the role that logical form must play if it is to have a role within philosophy, then it is unclear whether the linguistic conception of logical form can wholly supplant the traditional view. The linguistic conception of logical form seemingly has little to do with the conceptual analysis. And unless conceptual analysis takes the form of a grammatical analysis, it is unlikely that one can substitute grammatical analysis for the description of the logically significant aspects of our concepts.

This is not to deny that a linguistics-based conception of logical form is an important, maybe even essential part of understanding how to think about some aspects of logic and meaning. This is particularly clear with respect to the study of quantification. But there

are many questions about the nature of logical form that need to be resolved before particular view can be judged to be the most viable.

See also: Interpreted Logical Forms; Propositions; Quantifiers: Semantics.

Bibliography

- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky N (1977). 'On WH movement.' In Culicover P, Wasow T & Akamajian A (eds.) *Readings in English transformational grammar*. Waltham, MA: Ginn. 184–221.
- Chomsky N (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Davidson D (1967). 'Truth and Meaning.' *Synthese* 17, 304–323.
- Frege G (1952). *Translations from the philosophical writings of Gottlob Frege*. Oxford: Blackwell.

- Higginbotham J (1993). 'Logical form and grammatical form.' *Philosophical Perspectives* 7, 173–196.
- Huang C T J (1982). 'Move WH in a language without WH movement.' *Linguistic Review* 1, 369–416.
- Lasnik H (2001). 'Derivation and representation in generative grammar.' In Baltin M & Collins C (eds.) *Handbook of contemporary syntactic theory*. Oxford: Blackwell. 62–88.
- Lepore E & Ludwig K (2002). 'What is logical form?' In Preyer G & Peters G (eds.). 54–90.
- Ludlow P (2002). 'LF and natural logic.' In Preyer G & Peters G (eds.). 132–168.
- May R (1977). 'The grammar of quantification.' Ph.D. diss., MIT.
- Neale S (1994). 'Logical Form and LF.' In Otero C (ed.) *Noam Chomsky: critical assessments*. London: Routledge. 788–838.
- Preyer G & Peters G (eds.) (2002). *Logical form and language*. Oxford: Oxford University Press.
- Russell B (1919). *Introduction to mathematical philosophy*. London: George Allen and Unwin.
- Williams E (1983). 'Syntactic and Semantic Categories.' *Linguistics and Philosophy* 6, 423–446.

Lying, Honesty, and Promising

D Owens, University of Sheffield, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

Promising and asserting are both speech acts, and as such they are regulated by practical norms as well as linguistic norms (e.g., norms of etiquette). Here we shall be concerned with the **moral** norms that govern these speech acts.

Informational Theories

Many philosophers take the view that the morality of an act – its rightness or wrongness – is a function of the harms and benefits this act brings to the agent and to others. These philosophers also hold that the main way promises and assertions affect the interests of human beings is when they serve as sources of information. So, they conclude, both promises and assertions are morally significant principally because, and in so far as, they purport to offer information. Let's call this the 'informational' view of the morality of promise and assertion. On this view, morality censures an unfulfilled promise or a false assertion because these deeds can harm others by giving them false information.

We are all obliged to take due care not to lead others to form false beliefs, at least where this might be harmful to them (Scanlon, 1998: 300). This obligation means that we must not set out to deceive people by making them insincere promises or telling them things that we know to be false. But it also means that we mustn't change our minds about what we promised we were going to do (without good reason) or make an assertion without adequate evidence. Someone who accepts a promise standardly forms the expectation that the promisor will perform, and they may rely on this expectation to their detriment. Someone who believes an assertion is similarly exposed, if this assertion turns out to be false. I'll deal with informational theories of promising first and then move onto assertion.

Information theorists of promissory obligation fall into two categories. First, there are those (sometimes called 'expectation theorists') who argue that we are all under an obligation not to mislead others about how we shall behave in the future and that this obligation is why we ought not to make them promises that we do not fulfill (Scanlon, 1998: Chap. 7; Thomson, 1990: Chap. 12). Second, there are those who argue that we are obliged to fulfill our promises only where there is an up and running practice of fulfilling

one's promises: prior to this, there is no promissory obligation. For such 'practice theorists,' promissory obligation is conventional (Hume, 1978: Book III, Part II, Section V; Anscombe, 1981; Prichard, 1968).

Practice theories differ from expectation theories in their account of how the obligation to keep a promise arises from our interest in having correct information about how other people are going to behave. According to the practice theorist, we can create expectations of performance in our audience by uttering words like "I promise" only where there is an actual practice of making such utterances true, i.e., where people have come to feel some obligation to make them true. So one can't explain the moral significance of this utterance simply by reference to the expectations it creates. Still the practice theorist agrees with the expectation theorist that we are obliged to maintain the practice of promise-making where it exists, because this practice serves our information interest and thereby aids the coordination of behavior.

Turning now to assertion, there is much controversy among philosophers of language about the extent to which language in general, and assertion in particular, involve social convention. For example, Davidson maintains that "there is no known, agreed upon, publicly recognizable convention for making assertions" (Davidson, 1984: 270), and he thinks the same is true of promising. On the other hand Fried urges that the promisor has "intentionally invoked a convention whose function it is to give grounds – moral grounds – for another to expect the promised performance" and "to abuse that confidence is like . . . lying: the abuse of a shared social institution that is intended to invoke the bonds of trust" (Fried, 1980: 16). Thus, there is a division among information theorists of the morality of assertion parallel to that between expectation and practice theorists of promissory obligation.

It has long been debated whether there is any morally significant difference between lying to someone and deceiving them in a more oblique fashion (e.g., by way of false implicatures, deliberate ambiguity, or by leaving misleading evidence around, etc.). And this debate reflects a genuine ambivalence in our everyday attitudes. Where we feel entitled to deceive others – for the sake of their health for instance – many of us are still inclined to go to the trouble of trying to avoid telling a direct lie. On the other hand, where such deception is wrong, the wrong is seldom thought to be mitigated just because a direct lie was avoided.

There is a tradition of thought, however, according to which lying is always wrong, but we are sometimes permitted to deceive in other ways (Aquinas, 1966: II-II 110 a3; MacIntyre, 1995: 309–318). But many contemporary writers have expressed doubts about

whether the manner of the deception could by itself make a serious moral difference (Sidgwick, 1981: 354–355; (Williams, 2002: 100–110). An expectation theorist who maintains that the wrong of lying is only the wrong of deception will share these doubts (Scanlon, 1998: 320). On the other hand, a practice theorist of the morality of assertion may allow that, in addition to any harm he does to the person he deceives, the liar is abusing and thereby undermining a valuable social practice (Kant, 1991: 612), namely the use of language to convey information.

Noninformational Theories

Until now, we have been assuming that what makes an act, including a speech act, wrong is, some harm that it does to those it wrongs in the end. There are many moral theorists who reject this assumption, and it is open to them to propound noninformational theories of what is wrong with a lie or a broken promise. Rather than attempt a comprehensive classification of noninformational theories, I shall consider one such theory of promising and one such theory of lying, both taken from Kant.

Take promising. Kant locates the moral significance of promising not in the information interests it serves but rather in the fact that it grants the promisee a certain moral authority over the promisor: it entitles the promisee to require the promisor to perform and thus deprives the promisor of a certain moral freedom (Kant, 1996: 57–61). If I promise you a lift home, I am obliged to give you a lift unless you release me from this promise. This line of thought was central to classical theories of promissory obligation (Hobbes, 1991: Chap. 2) and has found an echo in some contemporary writing (Hart, 1967: 60). For these authors, a breach of promise wrongs the promisee, whether or not it also harms them by inducing false expectations in them, because it flouts the moral authority that the promisee has acquired over the promisor.

On this view, informational theories miss what is distinctive about promising. There are many ways of influencing people's expectations and of co-ordinating your behavior with others (Raz, 1977: 215–216). For example, one can predict that one will do something or even express a sincere intention to do something, while making it clear that one is not promising. To promise to do it is to express the intention to undertake an obligation to do it (Searle, 1969: 60), an obligation that mere expressions of intention or predictions do not bring down on the speaker, however firm or confident they may be. If I predict, on excellent evidence, that I shall be going in your direction because the police will be towing my car in

that direction, I have not committed myself to making it true that I shall be going in your direction when that prediction threatens to be falsified.

Turning now to lying, Kant makes a firm distinction between wrong one does in deceiving someone and the wrong one does by lying (Kant, 1996: 182–184). One can lie without deceiving (e.g., when one knows one won't be believed) and one can deceive without lying. On deceiving someone you may wrong them but, according to Kant, when you lie, the person you wrong is yourself. The liar violates a duty to himself (though this violation need not involve harming himself). On explaining the nature of this wrong, Kant follows thinkers like Aquinas in attributing a natural teleology to speech:

communication of one's thoughts to someone through words that yet (intentionally) contain the contrary of what the speaker thinks on the subject is an end that is directly opposed to the natural purposiveness of the speaker's capacity to communicate his thoughts (Kant, 1996: 182).

In lying, one violates a duty to oneself by abusing one's own faculties, by using oneself "as a mere means (a speaking machine)" (Kant, 1996: 183). It may be possible to capture Kant's basic idea here without reference to natural teleology or duties to self if we adopt a certain view of assertion.

One can distinguish two currently influential theories of assertion. In the first, inspired by Grice, asserting that *p* is a matter of uttering something with the intention of thereby getting your audience to believe that *p*, by means of their recognition of that very intention (Grice, 1989). If something like this statement is correct, the moral significance of an utterance *qua* assertion must lie solely in the effect it is trying to achieve, i.e., in the effect that the assertion has on the beliefs of others. In the second view of assertion, asserting that *p* is more like promising that *p*. In promising, someone intentionally undertakes an obligation to perform: undertaking such obligations is what promising consists in. Similarly to assert a certain proposition is, on this view, to intentionally undertake an obligation to ensure that one asserts only what is true (Dummett, 1973: 299–302) and, perhaps, to defend one's assertions as true, should they be challenged (Brandom, 1983). Putting oneself under such obligations is what assertion consists of.

Once the second view is in play, we can say what is wrong about lying without making reference to the effect that the lie has on others. A liar is in the

wrong not because he is wronging someone but because he knows that he is taking on obligations he cannot discharge. In this way, lying differs from deception that wrongs the deceived when it harms their interests in some way. Deception is an offence against others while lying is an offence against truth. Provided morality is not solely concerned with harm, this offence may be counted as a moral wrong.

Bibliography

- Aquinas T (1966). *Summa theologiae*. Gilby T (ed.). Cambridge: Blackfriars.
- Anscombe E (1981). *Ethics, religion and politics*. Oxford: Blackwell.
- Brandom R (1983). 'Assertion.' *Nous* 17(4), 637–650.
- Davidson D (1984). *Essays on truth and interpretation*. Oxford: Oxford University Press.
- Dummett M (1973). *Frege philosophy of language*. London: Duckworth.
- Fried C (1981). *Contract as promise*. Cambridge, MA: Harvard University Press.
- Grice P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Hart H (1967). 'Are there any natural rights?' In Quinton A (ed.) *Political philosophy*. Oxford: Oxford University Press. 53–66.
- Hobbes T (1991). *De cive*. Gert B (ed.). Indianapolis: Hackett.
- Hume D (1978). *Treatise on human nature*. Oxford: Oxford University Press.
- Kant I (1991). *Practical philosophy*. Gregor M (ed.). Cambridge: Cambridge University Press.
- Kant I (1996). *The metaphysics of morals*. Gregor M (ed.). Cambridge: Cambridge University Press.
- MacIntyre A (1995). 'Truthfulness, lies and moral philosophers.' In *Tanner Lectures on Human Values* 16. Salt Lake City: Utah University Press. 309–361.
- Prichard H (1968). *Moral obligation and duty and interest*. Oxford: Oxford University Press.
- Raz J (1977). 'Promises and obligations.' In Hacker P & Raz J (eds.) *Law, morality and society*. Oxford: Oxford University Press. 210–228.
- Searle J (1969). *Speech acts*. Cambridge: Cambridge University Press.
- Scanlon T (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Sidgwick H (1981). *The methods of ethics*. Indianapolis: Hackett.
- Thomson J (1990). *The realm of rights*. Cambridge, MA: Harvard University Press.
- Williams B (2002). *Truth and truthfulness*. Princeton: Princeton University Press.

This page intentionally left blank

M

Mass Nouns, Count Nouns, and Non-count Nouns: Philosophical Aspects

H Laycock, Queen's University, Kingston, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Linguists often distinguish ‘count nouns’ (CNs) and ‘non-count nouns’ (NCNs) (designated as ‘count+’ and ‘count’ nouns, respectively). The distinction, though hardly simple, is both exhaustive and entirely natural. In philosophical writings, by contrast, it is more usual to posit a dichotomy of ‘count nouns’ and ‘mass nouns’ (MNs) – a dichotomy that is very commonly (and however vaguely) supposed to be of metaphysical or ontological significance. But this dichotomy, unlike that of CNs and NCNs, is deeply problematic; here, in consequence, I speak only of a **supposed** dichotomy of CNs and MNs, and by the same token, of a **putative** category of MNs.

Plural Count Nouns and Non-count Nouns

There is a certain kinship between NCNs and plural CNs, a kinship that has in recent years attracted some attention; see e.g., Schein (1994). What is less widely appreciated is the basis of this kinship in the actual semantic status of NCNs. Laycock (1998) urged that since CNs, or their occurrences, are semantically **either** singular or plural, to be non-count is simply to be **neither** singular nor plural. Non-count nouns are then semantically non-singular, and it is this that underlies their kinship with the plural (plural nouns themselves, it is evident, are non-singular). But being non-plural, the non-count form is never, unlike that of many plural sentences, reducible to the canonical singular form. The relationships between the semantics of CNs and NCNs may thus be simply represented in **Table 1** (Laycock, 1998). The inclusion of a contrast between ‘clothes’ and ‘clothing’ alongside that between ‘apples’ and ‘water’ serves to emphasize the point that the singular/non-singular contrasts in general (and not simply the singular/plural contrasts) are first and foremost **semantic** as opposed to **metaphysical** or **ontic**

contrasts (it being assumed that the ‘clothes’/‘clothing’ contrast itself is purely a semantic one).

Non-count nouns are to be classed as semantically non-singular, simply by virtue of being non-count. And given this, both quantification and denoting that involve such nouns also must be semantically non-singular. In the case of quantification, the non-singularity of NCNs is reflected in their obvious non-acceptance of singular quantifiers – in the fact that we may speak (as with plural nouns) of *all water*, *some water*, and *more water* but not in the singular of *a water*, *each water*, or *one water*. *Any*, *all*, and *some* interact with NCNs much as with essentially plural nouns; and *All n V made of polyester* and *The n in the warehouse V made of polyester* are related in essentially the same way, whether *n* is replaced by *clothes* and *V* by *are*, or *n* is replaced by *clothing* and *V* by *is*. The implications of the non-singularity of NCNs for the semantics of sentences containing definite descriptions and referential expressions are complex and extend beyond the remit of these brief remarks. However, the relatively common **mereological** interpretations of non-count reference take it to be semantically singular, designating individual ‘parcels of stuff’ or ‘quantities’ (see e.g., Bunt, 1985), and the non-singularity of NCNs is evidently not consistent with approaches of this type. Indeed, difficulties with the singularity assumption have been noted precisely in relation to Russell’s theory of singular descriptions, which maintains, as he puts it, that *the* “in the singular” involves **uniqueness** (Russell, 1956: 176); see e.g., Montague (1973) and Laycock (1975).

The Concept ‘Mass Noun’ and Its Supposed Criterion

Turning now to so-called MNs, perhaps the first use of an expression of the MN genre occurs in Jespersen (1924), who wrote of ‘mass words,’ contrasting these with what he called ‘countables’ or ‘thing words.’ And the thought that such words had a distinct metaphysical significance received the following expression in his work:

Table 1 Semantics of count nouns versus non-count nouns

	1. Singular (‘one’)	2. Non-singular (‘not-one’)
3. Plural (‘many’)	– – –	‘things’ ‘apples’ ‘clothes’
4. Non-plural (‘not-many’)	‘thing’ ‘apple’ ‘piece of clothing’	‘stuff’ ‘water’ ‘clothing’

There are a great many words which do not call up the idea of some definite thing with a certain shape or precise limits. I call these ‘mass-words’; they may be either material, in which case they denote some substance in itself independent of form, such as ... water, butter, gas, air, etc., or else immaterial, such as ... success, tact, commonsense, and ... satisfaction, admiration, refinement, from verbs, or ... restlessness, justice, safety, constancy, from adjectives. (Jespersen, 1924: 198)

Subsequent writers typically differ from Jespersen in treating the domain of ‘mass words’ as one of **concrete** nouns exclusively; but insofar as these latter nouns are concerned, Jespersen’s approach would seem to represent the norm. Thus if the question is raised, of what semantic element constitutes the putative MNs as a distinct category of concrete nouns, the answer tends to be that it is precisely some such element of ‘form-independence’ – an absence of ‘criteria of distinctness’ (Hacker, 1980) or of a ‘boundary-drawing,’ ‘individuating,’ ‘form-specifying,’ or ‘reference-dividing’ component in their meaning (Quine, 1960; Chappell, 1971; Talmy, 1978; Langacker, 1987; Jackendoff, 1991; Kleiber, 1997).

In this regard, Quine nicely represented the common view. To learn a “full-fledged general term” like ‘apple’ it is not enough, so he remarks, to learn “how much of what goes on counts as apple”: “we must learn how much counts as **an** apple, and how much as another. Such terms possess built-in modes ... of dividing their reference” (1960: 91). So-called ‘mass terms,’ in contrast, do not thus divide their reference. Water, Quine wrote, “is scattered in discrete pools and glassfuls... still it is just ‘pool,’ ‘glassful,’ and ‘object,’ not ‘water’... that divide their reference” (1960: 91). If such a noun is used to individuate a full-fledged, ‘substantial’ object, it needs an individuating adjunct. There is no learning “how much counts as **some** water and how much counts as **some more**”; there is no such distinction to learn. Whereas any sum of parts that are each an apple is not another apple, this lack of a boundary-drawing element confers upon the putative MNs what Quine called “the semantical property of referring cumulatively”: “any

sum of parts which are water is water,” as he puts it. I shall call this widely accepted criterion for distinguishing the putative category of MNs from CNs, in whichever of the various equivalent forms it is fleshed out, the ‘no built-in reference-division’ (no-RD) criterion. The key assumption that underlies the supposed dichotomy of CNs and MNs is, then, that possible borderline cases apart, there is a specific and more or less determinate category of concrete nouns that answers to the no-RD criterion – the putative category of MNs, to be precise.

Now, whereas the range of nouns that are categorized as MNs varies significantly from one writer to another, all the nouns that are thus categorized are in fact (and must be) NCNs. But here my concern is not with the nouns themselves; it is just with the putative **category**, as determined by the no-RD criterion, to which they are said to belong. And insofar as the use of the term ‘MN’ rests on this criterion, the contrast of CNs and MNs is quite fundamentally misconceived; the reality is that no such category as that of MNs exists. There are, on the contrary, two semantically distinct categories of nouns that answer to the no-RD criterion – concrete NCNs and concrete plural CNs (a fact that, given the common non-singularity of these two categories, is not altogether surprising).

An Illusory Criterion

Although the kinship of the putative MNs with concrete plural CNs is commonly remarked, it is less often noted that the no-RD criterion itself applies identically to the plural nouns. Thus, for instance, although we learn “how much counts as **an** apple, and how much as another,” there is no learning “how much counts as **some** apples, and how much as **more** apples” – there is no such distinction to learn. Although the singular *apple* applies to just one apple at a time, *apples* sets no limits on the count of apples. It is not the meaning content of the plural noun itself that sets whatever limits there may be; it is contingencies of context, including acts of demonstration – for example, *these apples* – that demarcate the subject matter of a discourse. *Apples* provides no criteria of distinctness or boundaries for what it collectively applies to – it does not, *qua* plural, carve what it applies to ‘at the joints.’ To play the role of designating full-fledged objects, each of which is apples, *apples*, much like *water*, needs an individuating adjunct (‘heap of _____,’ ‘bag of _____,’ or the like). Thus, if water may be characterized as ‘form-indifferent,’ then apples, too, collectively, may be so characterized. Much as the water in a glass might be spilled or dispersed and survive, so, too, might the apples in a bag. And so far as Quine’s ‘cumulative

reference' is concerned, whereas any sum of parts each of which is **an** apple will not be **another** apple, any sum of parts that are **apples** will simply be more **apples**.

The appearance of a dichotomy between CNs and the putative MNs then arises purely and simply because the chosen occurrences of CNs are **singular** exclusively – plural nouns are nowhere in the picture – and once plural occurrences of CNs are factored in, the supposed dichotomy with CNs disappears. Insofar as the no-RD criterion is conceived as definitional of the distinctive status of some putative 'metaphysically interesting' class of nouns, this putative category of nouns is ill-defined or ill-conceived, and talk of such a category is best abandoned. The only categories that are to be legitimately contrasted with CNs are those of NCNs as such and of the various subcategories of NCNs. And although metaphysically interesting distinctions between CNs and some sub-groups of NCNs certainly do exist, such distinctions have nothing to do with the spurious category of MNs. They are not, that is, a function of the no-RD criterion, obtaining as they do between various groups of nouns, **all of which** satisfy that criterion. (It is then hardly surprising that what is supposed to count as an MN varies significantly from one sponsor of the concept to another: the no-RD criterion does zero work, and it is rather individual metaphysical intuitions that actually determine whether a given NCN is to be assigned to the putative category or not.) What the no-RD criterion reflects is simply the contrast between CNs in the singular, and non-singular nouns altogether generally, whether NCNs or plural CNs. But the central contrast in this domain is rather one between distinct forms of non-singularity – the plural, and the non-count – and as such, to repeat, this contrast is purely a semantic one.

The point retains its relevance at the formal level of the contrast between 'stuff' and 'things' themselves. These words are formally NCNs and CNs, respectively; but whereas the contrast of stuff and things is not infrequently treated by sponsors of the 'MN' category as if it were a metaphysical distinction, the fact is that 'stuff' may be and is applied to **things** without restriction ('the stuff in the basement' may just denote a pile of pots and pans, garden tools, old chairs, and bicycles).

The Non-metaphysical Goods

We are left, then, with an essentially semantic contrast between concrete CNs and NCNs. And what this semantic contrast embodies are distinct modalities for the determination and specification of amount or quantity. Count nouns embody one such modality – trivially, that of counting through the use of natural

number-related words – *one horse, so many things, too few clothes, a dozen eggs, a single professor*, etc.; and in this intuitive sense, counting is applicable to the denotata of CNs **exclusively**. Non-count nouns, by contrast, involve a form of what is naturally called 'measurement': *so much cotton, too much stuff, so little water, five tons of clothing*, etc. And although the denotata of NCNs may be **only** measured and not also counted, measurement as such is applicable to the denotata of both NCNs and CNs alike. We may, for instance, speak both of *75 ccs of water* and *75 ccs of poppy seeds*, both of *5.5 kilos of clothing* and *5.5 kilos of apples*. Furthermore, it seems clear that in contrast with counting, **any** real number can in principle be assigned to the measure of an amount of something. The concept of weight, for instance, is such that it is intelligible to assign a weight of n kilos (where n represents an integer), or of $n \times \pi$ kilos, to a quantity of snow (or rice, apples, clothing, underwear, water, etc.).

Intuitively, then, counting may be described as the determination of 'discrete' or 'discontinuous' quantity and measuring the determination of 'continuous' quantity. Of the two, discrete quantity seems privileged: there is exactly one non-relative way of determining the quantity of, say, eggs in a carton, which is precisely to count them. But there is no such unique way of determining, say, the quantity of cotton in a warehouse; this might be done, e.g., by volume, or by weight, or indeed by counting the number of bales; and these different measures cannot be expected to be correlated in any uniquely determinate way.

The contrast of discrete and continuous quantity is not directly ontological – it is not a matter of whether something consists of discrete 'bits' (visible or otherwise) or not. We may count planets, eggs, or horses to determine their number; we may weigh apples, snow, or clothing to determine their amount. The non-ontic nature of the contrast is perhaps especially striking in the juxtaposition of such words as the CN *clothes* (or, e.g., *boots and shoes*) and its cognate collective NCN *clothing* (or *footwear*). Although *clothing* represents continuous quantity and *clothes* discrete quantity, to say that **there is** clothing here or there is to say no more than that **there are** clothes here or there. In this respect, there is good sense in the remark of Quine: "The contrast lies in the terms and not in the stuff they name ... 'shoe' ... and 'footwear' range over exactly the same scattered stuff" (1960: 91).

The contrast 'lies in the terms': although there are **units** of clothing, furniture, etc. (individual pieces of clothing, pieces of furniture, etc.) – indeed, while collective nouns like *clothing* and *furniture* might be said to be ontologically **equivalent** to cognate CNs – such NCNs are no less **semantically** non-count than

non-collective nouns like *water* and *mashed potato*. Thus, although there is a straightforward sense to talk of the smallest **number** of clothes – a single item of clothing – there is no good sense to talk of the smallest **amount** of clothing – is one woolen winter coat the same amount of clothing as a single nylon stocking? In absolute terms, I've suggested, talk of amounts in relation to the denotata of NCNs (collective or otherwise) is simply ill-defined; and relative to some particular dimension such as weight or volume, there is no semantic rationale for specifying minimum amounts.

But although not directly ontological, the contrast of discrete and continuous quantity involves the **possibility** of certain ontic contrasts. Counting truisitically involves discrete units; and although what is measured may consist of discrete units, measurement as such does not require it, and there are ontic category-differences **within** the semantic category of NCNs. Thus, contrast the two groups of NCNs (a) *furniture, footwear, and clothing*, and (b) *rubble, sand, and snow*, with what may be called the 'pure' NCNs of group (c): *mashed potato, wine, and water*. The collective nouns of group (a) may be said to be 'object-involving,' in that they are semantically 'atomic' – there are units of furniture, footwear, and clothing that are not divisible into smaller units of furniture, footwear, and clothing. It is part of the **meaning** of such an NCN that like a typical CN, it ranges over discrete pieces, units, or elements of what the NCN denotes; indeed the very **identity** of some furniture is not to be distinguished from that of some pieces of furniture. For this reason, the identity of the denotata of group (a) nouns is independent of the identity of the materials of which those denotata are composed; some furniture can survive some loss of constituent materials – wood, cloth, stuffing, etc. – and remain the same (arguably, indeed, it is conceivable that all of the materials of some furniture could be replaced over time while the furniture retains its identity). But the same can hardly be said of the nouns in groups (b) and (c).

Now group (b), though not thus atomic, are object-involving in that they may be said to be 'semantically particulate': it is part of their meaning that what these words denote consists of discrete grains, flakes, bits, etc. – the difference being that the identity of some sand (snow, rubble, etc.) is **not** dependent on that of certain particular grains, flakes, or bits; it may be further crushed or pulverized and yet remain the same. In contrast with groups (a) and (b), however, no such object-involving concepts enter into the meanings of the group (c) terms. Whereas, for instance, to say that there is furniture or clothing in some region is to say that there are constituent pieces or units of furniture or clothing in that region, to say that there

is wine or mashed potato in some region is not to say that there are objects that can be characterized as 'pieces' or 'units' of wine or mashed potato in that region. In the nature of the case, there is here no comparable notion of a constituent piece or unit.

See also: Descriptions, Definite and Indefinite: Philosophical Aspects; Identity and Sameness: Philosophical Aspects; Interpreted Logical Forms; Logic and Language: Philosophical Aspects; Logical Form in Linguistics; Reference: Philosophical Theories.

Bibliography

- Bunt H (1985). *Mass terms and model-theoretic semantics*. Cambridge: Cambridge University Press, 1985.
- Cartwright H (1970). 'Quantities.' *The Philosophical Review* 79, 25–42.
- Chappell V C (1971). 'Stuff and things.' *Proceedings of the Aristotelian Society* 71, 61–76.
- Chellas B F (1979). 'Quantity and quantification.' In Pelletier (ed.). 227–231.
- Hacker P (1979). 'Substance: the constitution of reality.' In French P, Uehling T & Wettstein H (eds.) *Midwest studies in Philosophy* 4. Minneapolis: University of Minnesota Press. 239–261.
- Jackendoff R (1991). 'Parts and boundaries.' *Cognition* 41, 9–45.
- Jespersen O (1924). 'Mass-words.' In *The philosophy of grammar*. London: Allen and Unwin. 198–201.
- Kleiber G (1997). 'Massif / comptable et partie / tout.' *Verbum* 3, 321–337.
- Koslicki K (1999). 'The semantics of mass predicates.' *Nous* 33, 46–91.
- Koslicki K (2005). 'Nouns, Mass and Count.' In Borchert D M (ed.) *Encyclopedia of Philosophy*, 2nd edn. New York: MacMillan Reference.
- Langacker R (1991). 'Nouns and verbs.' In *Concept, image and the symbol*. Berlin: Mouton de Gruyter. Ch. 3.
- Laycock H (1975). 'Theories of matter.' *Synthese* 31, 411–442. [Reprinted with revisions in Pelletier (ed).]
- Laycock H (1998). 'Words without objects.' *Principia* 2(2), 147–182.
- Laycock H (2005). *Words without objects*. Oxford: Clarendon Press.
- Montague R (1973). 'The proper treatment of mass terms in English.' In Pelletier (ed.). 173–178. [Originally published (1973) as 'Response to Moravcsik.' In Hintikka K J J Moravcsik J M E & Suppes P (eds.) *Approaches to natural language*. Dordrecht: Reidel.]
- Parsons T (1979). 'An analysis of mass terms and amount terms.' In Pelletier (ed.). 138.
- Pelletier F J (ed.) (1979). *Mass terms, some philosophical problems*. Dordrecht: Reidel.
- Pelletier F J & Schubert L K (1989). 'Mass expressions.' In Gabbay D & Guenther F (eds.) *Handbook of Philosophical Logic 4: Topics in the philosophy of language*. Dordrecht: Reidel. 327–407.

- Quine W V (1960). *Word and object*. Cambridge, MA: MIT Press.
- Russell B (1956). *Introduction to mathematical philosophy*. London: Allen and Unwin.
- Schein B (1994). *Plurals and events*. Cambridge, MA: MIT Press.
- Sharvy R (1980). 'A more general theory of definite descriptions.' *The Philosophical Review* 89, 607–624.

- Strawson P (1953–1954). 'Particular and general.' *Proceedings of the Aristotelian Society* 54, 233–260.
- Talmy L (1978). 'The relation of grammar to cognition.' In Rudzka-Ostyn (ed.) *Topics in cognitive linguistics*. Amsterdam: John Benjamins. 165–207.
- Zimmerman D W (1995). 'Theories of masses and problems of constitution.' *The Philosophical Review* 104, 53–110.

Maxims and Flouting

A K Greenall, Norwegian University of Science and Technology, Trondheim, Norway

© 2006 Elsevier Ltd. All rights reserved.

Maxims and *flouting* are two closely interrelated terms central to the Oxford language philosopher H. Paul Grice's famous theory of the Cooperative Principle, which first emerged in the William James lectures that Grice delivered at Harvard University in 1967 (see **Cooperative Principle**). Grice's project was to try to reduce the number of meanings for lexical items ("Grice's razor" [Davis, 1998: 20]), and he did this by postulating a new, separate type of non-semantic meaning that he called *implicature*, a type of meaning that is not semantically coded but arises in conversational context (see **Implicature**).

The generation of implicature is crucially connected to the workings of an overall *Cooperative Principle* and a set of conversational *maxims*.

The Cooperative Principle

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged (Grice, 1989).

The Maxims

Quantity

1. Make your contribution as informative as is required.
2. Do not make your contribution more informative than is required.

Quality

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

Relation

1. Be relevant.

Manner

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief.
4. Be orderly.

These maxims may either be observed or breached, and in both cases implicatures may arise. The maxims can be contravened in a number of different ways, but crucial importance was allotted to the *blatant*, *intentional* contravention of one (or more of the) maxim(s), or *flouting*. On spotting such a blatant breach, the hearer – who will always retain his or her belief that the speaker is being cooperative – will immediately begin a search for an additional or alternative meaning for the utterance, one that *observes* the maxim(s) in question and thus follows the Cooperative Principle. The result of such a search is the implicature. Consider the following example:

- A: Are you going to Anna's party?
B: Well, Anna's got this new boyfriend now.

B's reply here clearly breaches Grice's maxim of Relation (the only truly relevant answers to A's question would be *yes* or *no*). According to the theory, spotting this breach would set off a reasoning process in A, along the following lines:

1. B has said that *Anna's got this new boyfriend now*.
2. This utterance breaches the maxim of Relation.
3. I have, nevertheless, no reason to believe that B does not intend to observe the Cooperative Principle and the maxims.
4. B could not be doing this unless what he really wanted to convey was something different from what he literally says.

5. On the basis of the available context (e.g., that B really liked Anna's old boyfriend), what B really wanted to convey was, *no* (*I am not going to Anna's party*) (a *relevant* reply, i.e., one that is not in breach of the maxim of Relation).

Note how this example illustrates the context-dependency of implicatures. If the context assumed in 5 were *B really disliked Anna's old boyfriend*, the interpretation (implicature) would be the opposite, namely, *yes* (*I am going to Anna's party*).

What Is a Maxim?

The concept of maxim is a crucial notion within the theory of the Cooperative Principle. Grice's own characterization of the entity is many-faceted. First of all, he was unambiguous on the point that the maxims are descriptive rather than prescriptive. Our rational nature, according to Grice, leads to the observable situation that the maxims are observed (more often than not). That is, he never meant that the maxims *should always* be observed, one of several common misunderstandings which has marred the field of Gricean pragmatics since its beginning (Thomas, 1995: 56).

Another noteworthy aspect of Grice's characterization of his maxims is that just because they are seen to have a basis in human rationality, they are not therefore to be considered innate. Furthermore, he was also attracted to the idea of maxims as *general interactional principles* governing both non-verbal and verbal behavior. Finally, the proposed set of maxims was seen as *expandable*: "There are, of course, all sorts of other maxims (aesthetic, social, or moral in character), such as 'Be polite,' that are also normally observed by participants in talk exchanges, and these may also generate . . . implicatures" (Grice, 1989: 28).

In the post-Gricean literature, there are roughly three main approaches to maxims. The first tries to tighten the association between the maxims and human rationality. The main representative of this approach, Asa Kasher, does this by postulating principles of rationality from which Grice's maxims are seen to derive (Kasher, 1976). The second, more influential approach, alienates itself most strongly from Grice's original formulation: here, the maxims are redefined as innate, linguistic rules that are always observed and that will produce the same implicatures irrespective of context (e.g., Atlas and Levinson, 1981; Horn, 1984). This formalist, cognitivist approach, in its quest to 'purge' the Gricean scheme, often ends up reducing the number of proposed maxims, a trend that culminated in the theory of relevance proposed by Sperber and Wilson in 1986, in which all the maxims were collapsed into one, cognitive,

Principle of Relevance. This is in stark contrast to a completely opposite trend, found in the third approach to maxims, the social pragmatics approach. Geoffrey Leech (1983), adopting Grice's view of maxims as learned entities, ends up adding a large number of maxims to Grice's scheme.

The main drawback of the formalist, cognitivist approach is clear: they see maxims as rules that are in principle always followed, and hence this kind of theory cannot account for that which is observed when some such entity as a maxim is *breached* (e.g., flouted), namely, the emergence of a layer of underlying meaning (implicature). Support for this observation has come from a perhaps unusual angle: research in Artificial Intelligence has shown that, in order for computers to communicate efficiently with human beings, they need to observe all the Gricean maxims at all times. If they do not, the human communicator will invariably read more into the computer's utterance than the devisers of the program intended (Bernsen *et al.*, 1996).

The main criticism of the social pragmatic approach has revolved around their uncritical, *ad hoc* proliferation of maxims (see, e.g., Brown and Levinson, 1987). A weakness of *all* approaches, including the social pragmatics approach, is the widespread belief that there must be an ultimate, fixed set of universal maxims where each can be classified and labeled individually. This despite the fact that the universality of maxims was drawn strongly into question as early as the mid-1970s, in Elinor Keenan's (1976) famous study of a Malagasy community, in which the influence of an informativity maxim (Quantity) was shown to be absent in certain speech situations. Unfortunately, rather than drawing attention to the eclectic sociocultural reality of Gricean maxims and other such entities, Keenan's work was applauded as a refutation of Grice's claims by those who believe that universality is the only valid stamp of approval for maxims.

The varied appearance of maxims around the world makes likely the hypothesis that one cannot ever hope to be able to enumerate and label every single maxim – the list of entities that can be flouted to produce implicature is probably open-ended. This, of course, calls for a detailed and succinct definition of the notion, to stand guard at the door, controlling which new maxims get in. A simple solution would be to adopt a functional definition such as that proposed by François Recanati (1987: 133), to the effect that a maxim is an entity that can be flouted to produce implicature (see also Levinson's 'general principle' [1983: 132]). The success of such a definition would, however, depend on whether or not one has reached a satisfactory understanding of the notion of *flouting*.

Flouting: Past, Present, and Future

As was noted in the first section of this article, flouting is the blatant breach of one of the maxims. Because all hearers embody faith in the speaker's inherent intention to be cooperative (i.e., to observe the maxims [at some level]), a seeming breach will trigger a reasoning process whereby the hearer will try to come up with a meaning for the utterance that *turns it into* an act of observing the given maxim(s) (an implicature).

There are several problems with this model. First of all, it is unlikely that it is the hearer's belief in the speaker's cooperativity that is the reason behind the former's quest to retrieve an implicature. This is especially the case if one chooses to include politeness maxims into the scheme: if a politeness maxim is breached, the implicature is hardly ever more polite than the act itself (consider, e.g., the act of turning away when one should have greeted somebody, in order to implicate, for instance, that one considers that person to have done something despicable). Another serious problem is that the reasoning process the hearer is supposed to go through is psychologically implausible. And finally, the end product of this process, the implicature, is presented as an entity that is easily circumscribed (a fully fledged proposition), whereas real life might prove to be far less ordered, with several half-formed, semi-verbalized hypotheses as outcome.

Despite these problems, which are clearly in need of serious attention, there has been a curious neglect of the notion of flouting in post-Gricean theory-building. The interest from empirical researchers has, by contrast, been tremendous. The notion is used as an analytical tool in a large array of studies in a multitude of extra-pragmatic and semi-pragmatic fields such as, e.g., literary criticism (e.g., Bollobás, 1981), humor studies (e.g., Attardo, 1990), or gender studies (Rundquist, 1992). Such widespread application clearly confirms the theoretical value of the notion. However, as very few have felt the need to make any significant amendments to the theory (apart from the occasional addition of a 'new maxim'), the problems remain.

In the theoretical literature, it is mainly the strong focus on observance (largely because of a reluctance to deal with the troublesome notion of context) that has obscured the need to deal with the notion of flouting. The discussion above related the formalist approach with their definition of maxims as context-independent rules that are always to be observed. Another trend manifests itself by interpreting every observed breach of a maxim not as a breach, but as the observance of an invented, 'opposite' maxim

(e.g., a breach of Be brief would be seen as an observation of 'Be wordy') (see, e.g., Joshi, 1982).

A particularly serious form of neglect is, of course, outright rejection, as in Sperber and Wilson's highly influential, and heavily criticized, Relevance Theory (1995). (see **Relevance Theory**.) Maxims, as envisaged by Grice, are afforded no place in a theory of indirectness, and the notion of flouting leading to implicature is replaced by the idea that implicature should be arrived at by a continued search for relevance, when the literal interpretation yields none. The problem for Relevance Theory is, of course, empirical evidence such as that mentioned in the previous section (Bernsen *et al.*, 1996), which shows without a shadow of a doubt that maxims (or something like them) *are* involved and that breaching them *does* have the effect of stimulating the interpretation of indirect meaning. In addition, there are interesting indications that Grice may have stumbled on to something that is in reality part of an even larger scheme than he envisaged. In experimental social psychology, it has long been noted that 'unusual events' are sociocognitive triggers for a search for an 'explanation' (the 'expectancy principle' [Weiner, 1985: 81]). The breach of a maxim, being the departure from a norm, is also in this sense an 'unusual event,' and the search for an implicature has much in common with the description of the search for an 'explanation.' If this connection is viable, it holds great promise for the further development of Grice's model of flouting.

Conclusion

All in all, Grice himself offered merely 20 or so pages on the topic of maxims and flouting. Subsequent research offers much food for thought, but no viable theory has emerged that exploits the full potential of Grice's eminent point of departure. This does not mean that it is an impossible task to arrive at such a theory, but it would probably require a considerable reorientation of focus, possibly toward the sociocultural significance of both maxims and flouting.

See also: Cooperative Principle; Implicature; Irony; Relevance Theory.

Bibliography

- Atlas J D & Levinson S C (1981). 'It-clefts, informativeness and logical form: radical pragmatics.' In Cole P (ed.) *Radical pragmatics*. New York: Academic Press. 1–16.
- Attardo S (1990). 'The violation of Grice's maxims in jokes.' *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, 355–362.

- Bernsen N O, Dybkjær H & Dybkjær L (1996). 'Cooperativity in human-machine and human-human spoken dialogue.' *Discourse Processes* 21, 213–236.
- Bollobás E (1981). 'Who's afraid of irony? an analysis of uncooperative behavior in Edward Albee's *Who's afraid of Virginia Woolf?*' *Journal of Pragmatics* 5, 323–334.
- Brown P & Levinson S C (1987). *Politeness: some universals in language usage* (2nd edn.). Cambridge: Cambridge University Press.
- Davis W A (1998). *Implicature: intention, convention, and principle in the failure of Gricean theory*. Cambridge: Cambridge University Press.
- Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Horn L R (1984). 'Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature.' In Schiffrin D (ed.) *Georgetown University Round Table on Languages and Linguistics 1984*. Washington, DC: Georgetown University Press. 11–42.
- Joshi A K (1982). 'Mutual beliefs in question-answer systems.' In Smith N V (ed.) *Mutual knowledge*. London: Academic Press. 181–197.
- Kasher A (1976). 'Conversational maxims and rationality.' In Kasher A (ed.) *Language in focus: foundations, methods and systems*. Dordrecht: D. Reidel Publishing Company. 197–216.
- Keenan E O (1976). 'The universality of conversational postulates.' *Language in Society* 5, 67–80.
- Leech G (1983). *Principles of pragmatics*. London: Longman.
- Levinson S C (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Recanati F (1987). *Meaning and force: the pragmatics of performative utterances*. Cambridge: Cambridge University Press.
- Rundquist S (1992). 'Indirectness: a gender study of flouting Grice's maxims.' *Journal of Pragmatics* 18, 431–449.
- Sperber D & Wilson D (1995). *Relevance: communication and cognition* (2nd edn.). Oxford: Blackwell.
- Thomas J (1995). *Meaning in interaction*. London: Longman.
- Weiner B (1985). "'Spontaneous" causal thinking.' *Psychological Bulletin* 97, 74–84.

Meaning: Cognitive Dependency of Lexical Meaning

P A M Seuren, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

It is often thought or implicitly assumed, even in circles of professional semanticists, that predicate meanings, as codified in their satisfaction conditions, are lexically fixed in such a way that they automatically produce truth or falsity when applied to appropriate reference objects. This assumption is unfounded. In many, perhaps most, cases, the satisfaction conditions imply an appeal to nonlinguistic knowledge, so that the truth and falsity of assertive utterances are not the product of mere linguistic compositional computation, but are codetermined by nonlinguistic knowledge, either of a general encyclopedic or of a context-bound, situational nature.

An obvious case is provided by a large class of gradable adjectival predicates, such as *expensive*, *old*, and *large*, whose applicability depends on (preferably socially recognized) standards of cost, age, and size, respectively, for the objects denoted by their subject terms. The description of such standards is not part of the description of the language concerned, but of (socially shared) knowledge.

Further obvious examples are 'possession' predicates, such as English *have*, *lack*, and *with(out)*, and whatever lexical specification is needed for genitives,

datives, and possessive pronouns. These clearly require general encyclopedic knowledge for their proper interpretation. Consider the following examples:

- (1a) This hotel room has a bathroom.
- (1b) This student has a supervisor.

For (1a) to be true, it is necessary that there be one unique bathroom directly connected with the room in question, whose use is reserved for the occupants of that room. When the room carries a notice that its bathroom is at the end of the corridor to the right, while the same bathroom serves all the other rooms in the corridor, (1a) is false – not just misleading but false, as any judge presiding over a court case brought by a dissatisfied customer will agree. But for (1b) to be true, no such uniqueness relation is required, as one supervisor may have many students to look after. This is not a question of knowing English, but of knowing about the world as it happens to be.

The same goes for the parallel sentences:

- (2a) This is a hotel room with a bathroom.
- (2b) This is a student with a supervisor.

Possession predicates, therefore, must be specified in the lexicon as involving an appeal to what is normally the case regarding their term referents. They express a well-known relation of appurtenance between the kind of object referred to in subject position and the kind of object referred to in object position.

The semantic description (satisfaction conditions) of *have* and other possessive predicates is thus taken to contain a parameter for 'what is well-known,' making the interpretation of this predicate in each token occurrence truth-conditionally dependent on world knowledge.

Not all possession predicates are subject to the same conditions. Possessive pronouns, for example, may express a relation of 'being responsible for' or 'taking care of,' which other possession predicates cannot express. An example is sentence (3) uttered by a gardener with regard to the flower beds he is tending:

(3) Please don't mess up my flower beds.

This sentence can be uttered appropriately without the speaker implying that the flower beds are owned by him.

Many such examples can be given. Consider the predicate *flat* said of a road, a tire, a mountain, a face, or the world. There is an overall element 'spread out, preferably horizontally, without too much in the way of protrusions or elevations,' but that in itself is insufficient to determine what 'being flat' amounts to in these cases. The full meaning comes across only if it is known what roads, tires, mountains, faces, and the world are normally thought to be like. Dictionaries, even the best ones, limit themselves to giving examples, hoping that the user will get the hint.

Another example is the predicate *fond of*, as in:

(4a) John is fond of his dog.

(4b) John is fond of cherries.

(4c) John is fond of mice.

In (4a), obviously, John's fondness is of a rather different nature from what is found in (4b): the fondness expressed in the one is clearly incompatible with the fondness expressed in the other. The fondness of (4c) can be either of the kind expressed in (4a) or of the kind expressed in (4b). The common element in the status assigned to the object-term referents is something like 'being the object of one's affection or of one's pleasure,' but again, such a condition is insufficient to determine full interpretation.

Cognitive dependency is an essential aspect in the description of predicate meanings. The fact that some predicate meanings contain a parameter referring to an available nonlinguistic but language-independent, cognitive knowledge base means that neither utterance-token interpretation nor sentence-type meaning is compositional in the accepted sense of being derivable by (model-theoretic) computation from the linguistic elements alone. As regards utterance-token interpretation, this is already widely accepted, owing to valuable work done in

pragmatics. The noncompositionality of sentence-type meaning, defined at the level of language description, is likewise beginning to be accepted by theorists of natural language. This type-level noncompositionality does not mean, however, that the specification of the satisfaction conditions of predicates is not truth-conditional, only that standards embodied in socially accepted knowledge have become part of the truth conditions of sentences in which the predicate occurs.

In most treatises on lexicology, the term polysemy is used for phenomena such as those presented above. At the same time, however, it is widely recognized that this is, in fact, little more than a term used to give the problem a name. The problem itself lies in the psychology of concepts. One may assume that there are socially shared concepts like 'possession,' 'flatness,' and 'fondness,' but it is not known in what terms such concepts are to be defined. In a general sense, Fodor (1975, 1998) is probably right in insisting that lexical meanings are direct reflexes of concepts that have their abode in cognition but outside language. The necessary and sufficient conditions taken to define the corresponding lexical meanings cannot, according to Fodor, be formulated in natural language terms, but must be formulated in a 'language of thought,' which is categorically different from any natural language and whose terms and combinatorial properties will have to be established as a result of psychological theorizing.

It is clear, in any case, that phenomena like those shown in (1)–(4) pose a serious threat to any attempt at setting up a model-theoretic theory of lexical meaning, such as Dowty (1979): the neglect of the cognitive factor quickly becomes fatal in lexical semantics.

Context-bound or situational knowledge plays a role in the interpretation of predicates that involve a 'view-point' or 'perspective,' such as the pair *come* and *go*, or predicates such as *to the right (left) of*, *in front of*, and *behind*. The two versions of (5) are truth-conditionally identical, but they differ semantically in that the 'mental camera,' so to speak, has stayed in the corridor in the *went* version, but has moved along with Dick into the office in the *came* version.

(5) Dick and Harry were waiting in the corridor.

Then Dick was called into the office. After five minutes, Harry [went/came] in too.

In similar manner, the sentences (6a) and (6b) may describe the same situation, but from different points of view. In (6a), schematically speaking, the viewer, the tree, and the statue are in a straight line; in (6b), it is the viewer, the tree, and the fountain that are in a straight line:

- (6a) There was a statue behind the tree, and a fountain to the left of the tree.
- (6b) There was a fountain behind the tree, and a statue to the right of the tree.

A further cognitive criterion for the lexical meaning of predicates, especially those denoting artifacts, seems to be the function of the objects denoted. What defines a table or a chair is not their physical shape or the material they are made of, but their socially recognized function. The same holds for a concept like 'luxury.' Laws imposing special taxation on luxury goods or luxury activities usually enumerate the goods and activities in question, making exceptions for special cases (such as frock coats for undertakers). Yet what defines luxury is not a list of goods or activities, but socially recognized function – roughly, anything relatively expensive and exceeding the necessities of life.

A peculiar example of cognitive dependency, probably based on function, is provided by the English noun *threshold* and its Standard German translation equivalent *Schwelle*. In their normal uses, they denote the ridge or sill usually found between doorposts at floor level. Yet these two words differ in their capacity for semantic extension: the elevations in roads and streets that are normally called *speed bumps* in English are called *Schwelle* in German. Yet it is

unthinkable that speed bumps should be called *thresholds* in English. The question is: why? One is inclined to think that, at some ill-understood level of interpretation, the word *threshold* implies containment within a space or a transition from one kind of space to another, perhaps as a result of its etymology (which is not fully known). *Schwelle*, by contrast, is a swelling in the ground that forms an obstacle to be got over – which is also its etymology, although, on the whole, German speakers do not realize that. The difference between the two words is not a question of the ontological properties of the objects concerned, but, apparently, of the ways they are conceived of. The role of etymology in this case is intriguing.

See also: Polysemy and Homonymy.

Bibliography

- Dowty D (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Fodor J A (1975). *The language of thought*. Hassocks, Sussex: Harvester Press.
- Fodor J A (1998). *Concepts: Where cognitive science went wrong*. New York: Oxford University Press.

Meaning: Development

E V Clark, Stanford University, Stanford, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

How do children assign meanings to words? This task is central to the acquisition of a language: words allow for the expression of the speaker's intentions, they combine to form larger constructions, and the conventional meanings they have license their use for making references in context. Without them, there is no language. In the acquisition of meaning, children must solve the general mapping problem of how to line up word forms with word meanings. The forms are the words they hear from other (mainly adult) speakers. The meanings they must discern in part from consistencies in speaker usage in context from one occasion to the next and in part from inferences licensed by the speaker on each occasion. Possible meanings for unfamiliar words, then, are built up partly from children's conceptual representations of events and partly from the social interactions at the heart of adult-child conversation.

One critical task for children is that of working out the conventional meanings of individual words (e.g., *cup*, *team*, *friend*, *truth*). Yet, doing so is not enough: syntactic constructions also carry meanings that combine with the meanings contributed by the actual words used (causative constructions, as in *They broke the cup* or *The boy made the pony jump*; the locative construction, as in *She put the carving on the shelf*; the resultative construction, as in *He washed the floor clean*). However, children start mapping word meanings before they begin combining words.

Languages differ in how they lexicalize information – how they combine particular elements of meaning into words – and in the kinds of grammatical information that have to be expressed. They may package information about events differently; for example, combining motion and direction in a single word (*depart*) or not (*go + toward*), combining motion and manner (*stroll*), or not (*walk slowly*). They also differ in the grammatical distinctions made in each utterance. Some always indicate

whether an activity was completed; others leave that to be inferred. Some always indicate whether the speaker is reporting from direct observation, or, for example, from the report of someone else. Some indicate whether object-properties are inherent or temporary. The grammatical distinctions that languages draw on vary, as do the ways in which they lexicalize information about objects and events. Mapping meanings onto words is not simply a matter of equating meanings with conceptual categories. Children have to select and organize conceptual information as they work out what the conventional meanings are for the words they are learning.

How do children arrive at the meanings they first assign to unfamiliar words? How do they identify their intended referents? And how do they arrive at the relations that link word meanings in different ways? The general conversational context itself serves to identify relevant information on each occasion for children trying to work out the meaning of an unfamiliar word. Adult language use presents them with critical information about how words are used, their conventional meanings, and the connections among words in particular domains.

Conventionality and Contrast

Adult speakers observe two general pragmatic principles when they converse. First, they adhere to the conventions of the language they are speaking and in so doing make sure their addressees identify the meanings intended in their utterances. The principle of conventionality takes the following form: 'For certain meanings, there is a form that speakers expect to be used in the language community.' So if there is a conventional term that means what the speaker wishes to convey, that is the term to use. If the speaker fails to use it or uses it in an unusual way, that speaker risks being misunderstood. For conventions to be effective, conventional meanings must be given priority over any nonconventional ones. The second general principle speakers observe is that of contrast: 'Speakers take every difference in form to mark a difference in meaning.' When speakers choose a word, they do so for a reason, so any change in word choice means they are expressing a different meaning.

These two principles work hand-in-hand with the Cooperative principle in conversation and its attendant maxims of quality (be truthful), quantity (be as informative as required), relation (make your contribution relevant), and manner (avoid ambiguity; Grice, 1989). Acting in a cooperative manner demands that one observe the conventions of the language in order to be understood. At the same time, if there is no conventional term available for the meaning to be

expressed, speakers can coin one, provided they do so in such a way that the addressee will be able to interpret the coinage as intended (Clark, 1993).

In Conversation

Adults talk to young children from the very start, and what they say is usually tied closely to specific objects and activities. This feature of conversation presents young infants with opportunities to discern different intentions, marked by different utterances from early on. Infants attend to adult intentions and goals as early as 12 months of age. They show this, for example, by tracking adult gaze and adult pointing toward objects (e.g., Carpenter *et al.*, 1998), so if they are also attentive to the routine words and phrases used on each type of occasion, they have a starting point for discerning rational choices among contrasting terms and gestures.

Consider the general conditions for conversational exchange: joint attention, physical co-presence, and conversational co-presence. Adults observe these conditions and indeed impose them, as they talk to very young children. They work to get 1- and 2-year-olds to attend, for instance when planning to tell them about an unfamiliar object, and only then do they talk to them about whatever object or event is visibly present (Clark, 2001). By first establishing joint attention, adults set young children up to identify and then to help add to common ground. Children can do this by ratifying offers of new words by repeating them or else indicating in some other way that they have taken up an unfamiliar term (Clark, 2003).

When adults offer unfamiliar words, they do so in the conversational context; that is, with children who are already attending to whatever is in the locus of joint attention. This feature, along with any familiar terms that are co-present in the conversation, allows children to make a preliminary mapping by identifying the intended referent, whether it is an object or an action (Tomasello, 2002). In effect, the conditions on conversation narrow down the possible meanings that young children might consider for a new term to whatever is in the current joint focus of attention.

However, adults do more in conversation. They accompany their offers of unfamiliar words with additional information about the intended referent on that occasion and about how the target word is related to other terms in the same semantic field. Among the semantic relations adults commonly offer are inclusion (*An X is a kind of Y*), meronymy or partonymy (*An X is part of Y*), possession (*X belongs to Y*), and function (*X is used for Y*; Clark and Wong, 2002). After offering one term, adults often offer others that happen to contrast in

that context, so a dimensional term like *tall* may be followed up by *short*, *wide*, *narrow*, and *long* (Rogers, 1978). In fact, the meanings of words for unfamiliar actions may also be inferred in part from their co-occurrence with terms for familiar objects affected by those actions, and the meanings of words for unfamiliar objects may be inferred in part from the verbs with which the nouns in question occur (e.g., Goodman *et al.*, 1998; Bowerman, 2005). All this information offers ways for children to link new terms to any relevant words they already know. Children learn from child-directed speech about general properties of the lexicon – taxonomic relations, nonoverlapping categories within levels, opposites, overlaps in meaning (through hierarchical connections) vs. in reference, and so on.

In short, adults are the experts in providing the conventional terms used for specific meanings in the speech community. The novices, children, ask them innumerable *What's that?* questions from around age 2;0–2;6 on and treat them as reliable sources for how to talk about new things (e.g., Diesendruck and Markson, 2001). Moreover, when young children make errors, adults frequently check up, through side sequences and embedded corrections, on what they intended to say, and so present children with the conventional forms to aim for (Chouinard and Clark, 2003).

Making Inferences

When children hear a new term for some object or activity, they can infer in context that the term probably applies to the object or activity to which they are attending. However, the information that adults often follow up with allows children to make more detailed inferences about the candidate meaning. Mention of class membership – for example, *A sparrow is a bird* – tells them that they can add the term *sparrow* to the set of terms they already know for birds, perhaps just *chicken* and *duck*. Comments on the size, characteristic song, or flight each allow further inferences about how sparrows differ from ducks and chickens.

What evidence is there that young children take in such information? In spontaneous conversations, they give evidence of attending to what adults say in several ways. First, they repeat new terms in their next conversational turn, either as single words or embedded in a larger utterance; second, they acknowledge the adult offer with forms like *yeah*, *uh-huh*, and *mmh*; and third, they continue to talk about the relevant semantic domain (Clark, 2004).

Children's readiness to make inferences from added information offered by adults has also been examined

in word-learning experiments. In one study, children aged just 2 years old were taught words for two sets of objects (A and B) that were similar in appearance and had the same function. After teaching the first word for the first set (A), the experimenter introduced the second set of objects while saying just once, "Bs are a kind of A." He then proceeded to teach the second word, B. Children were then tested by asking them to find all the As and then all the Bs. For the first request, they typically selected As; for the second, they consistently (and correctly) picked only Bs (Clark and Grossman, 1998). In short, the one statement of an inclusion relation was enough for even young 2-year-olds to make use of it in this task. In another condition, again teaching two new words for two sets that resembled each other, children infer that there could be an inclusion relation but they have no way to tell which way it should go, so some include A in B, and some B in A.

Children rely on contrast in context to make inferences about the most probable reference for a newly introduced word. For example, if they already know what the object they are attending to is called, they are more likely to infer that a new term denotes a subordinate, a part, or some other property of it (Taylor and Gelman, 1989). This propensity was exploited directly in studies of whether children could decide in context whether a new word was intended to denote an object or an activity. Young 2-year-olds were presented with the same object doing different actions, with one action labeled with the new term, or else several objects, one labeled with the new term and all doing the same action. The children readily inferred that the new word denoted an activity in the first case and an object in the second (e.g., Tomasello, 2002).

Young children are also able to discern the intended from the accidental. When shown various actions being demonstrated, infants aged 18 months imitated intended actions (marked by utterances like *'There'*) more frequently than unintended ones (signaled by utterances like *'Oops'*). By age 2, young children know to ignore errors in wording, for example, and attend only to the final formulation of what someone is saying. In one study, for example, children were taught a word for a set of objects; then the experimenter exclaimed, "Oh, I made a mistake: these aren't As, they're Bs" and proceeded to teach the word B in place of the earlier A. When tested, even children who were just 2 years old knew that they did not know what A, the first word, meant (e.g., Clark and Grossman, 1998).

All the inferences presented so far have been overt inferences about unfamiliar word meanings, made on the spot by children exposed to the new words.

Yet, although adults make clear offers of new words, marking them as new by introducing them in formulaic deictic frames (e.g., *This is a . . .*), with utterance-final stress, many of the other words they use will be unfamiliar to very young children. How do children assign meanings to all those words? The answer lies in the covert use of Roger Brown's (1958) "original word game." Basically, the child notices an unfamiliar word, makes inferences in context about its probable meaning and acts on that, and then adjusts those inferences in light of the adult's responses. Consider these scenarios by way of illustration:

- (a) Young child watching parent in the kitchen, with several drink containers on the counter
 Mother (to older sibling): Hand me that mug, will you?
 (Child, wondering what a mug is, watches sibling pick up a mug)
 Mother: Thanks
 (Child infers for now that *mug* denotes something that has a handle, is a solid color, and is made of ceramic)

Sometimes, the inferences that children make are informed slightly more directly by the parent's direct responses, as in (b).

- (b) Young child holding two plastic animals, a cat and a dog
 Father: Can you give me the spaniel?
 (Child, uncertain what spaniel means, holds out the cat)
 Father: No, the spaniel please.
 (Child infers that *spaniel* must refer to a kind of dog rather than a kind of cat, and so hands over the plastic dog instead)

In both cases, the child makes preliminary or tentative inferences that can then be adjusted or changed in light of adult follow-up utterances, further exposures in other contexts, and additional, often explicit information about inclusion, parts, properties, or functions. Of course, inferences like these can also be made about terms for actions, relations, and states, as well as about those for objects, parts, and properties.

Pragmatics and Meaning

In the conversational exchanges considered so far, adult and child both follow the cooperative principle characterized by Grice (1989), as seen by their observation of joint attention, physical co-presence, and conversational co-presence. In addition, each participant in the exchange must add to common ground and keep account of the common ground that has been accumulated so far (H. Clark, 1996). All of this requires that speakers keep careful track of

the intentions and goals being conveyed within an exchange (Tomasello, 1995; Bloom, 2000).

Infants are attentive to nonlinguistic goals very early. For example, if 14-month-olds are shown an unusual action that achieves a goal – for example, an adult bending down to touch a panel switch with her forehead – they imitate it. If 18-month-olds watch an adult try and fail to place a metal hoop on a prong, the infants will produce the action successfully, even though they have never seen it completed (Meltzoff, 1995). That is, infants infer that the adult intended to turn on the light or intended to hang up the hoop. Intentions are what is critical, Meltzoff demonstrated, not just observation of the relevant actions, because infants do not re-enact these events when the actions are performed by a mechanical hand.

In much the same way, infants attend to the words that adults use. Upon hearing a word, they infer that the speaker is referring to the entity physically present in the locus of joint attention. If instead the speaker produces a different word, they infer that the speaker is now referring to something else and therefore has a different goal in speaking. That is, each linguistic expression chosen indexes a different intention, thus exemplifying the speaker's reliance on contrast, as well as on conventionality (Clark, 1993). This recognition then licenses young children to use words to express their intentions and in this way to convey specific goals. Adult usage provides the model for how to do so within conversational exchanges.

Infants also grasp quite early that the words used to express certain meanings are fixed and conventional. For example, they know that adults who wish to refer to a squirrel use the term 'squirrel' or to refer to a sycamore tree use the term 'sycamore,' and so on. As a result, when they notice adults who fail to use 'squirrel' when looking at a squirrel, but instead use another expression, they can readily infer that the speaker must therefore mean something else. In effect, young children, just like adults, assume that if the speaker intends to talk about a squirrel, he will use the conventional term for it. If instead, he uses something else, then he must intend to convey some other meaning.

As a result, in situations where children already know terms for some of the objects they can see, they expect the adult to use a familiar term for any familiar object. If the adult instead produces an unfamiliar term, in the presence of an unfamiliar object, they will infer that he intended to refer to the object for which they do not yet have a term. So they use contrast, together with the assumption that conventional terms always take priority, to interpret the speaker's intentions on such occasions. The result is that they consistently assign unfamiliar terms to as-yet unnamed objects or actions. This pragmatic

strategy for interpreting intentions and thereby making a first assignment of meaning to an unfamiliar word helps young children in many settings. Take the case of an adult looking at a single familiar object that is well known to the child. The adult, clearly talking about that object, does not use the expected conventional term. What can the child infer from that? There are two common options: (1) the unfamiliar expression denotes a superordinate or subordinate category, or (2) it denotes a part or property of the familiar object. Then, the remainder of the utterance can typically provide the child with important clues about the correct inference. For example, production of a familiar term for a known object is typically followed by a part term accompanied by a possessive pronoun (as in *his ear*), whereas such expressions as *is a kind of* or *is a* are associated with assignments to class membership in a superordinate category (Clark and Wong, 2002; Saylor and Sabbagh, 2004). Use of a verb like *looks* or *feels* (as in *it looks smooth, it feels soft*) often accompanies the introduction of property terms, and when the unfamiliar term is introduced before the familiar one with *kind of* (*a spaniel is a kind of dog*), the child readily infers that the new term, here *spaniel*, designates a subordinate category. Finally, children as young as age 2 rely on a combination of syntax cues and physical co-presence in identifying generic noun phrases; for example, when asked something like *What noise do dogs make?* with just one dog in sight.

What these findings indicate is that even very young children are highly attentive to the locus of joint attention and to whatever is co-present physically and conversationally. When one adds in whatever linguistic knowledge children have already built up about word meanings and constructions, it becomes clear that they have an extensive base from which to make inferences about possible, plausible meanings of unfamiliar words. This holds whether the words are presented explicitly as ‘new’ by adult speakers or whether children simply flag them *en passant* as unfamiliar and therefore in need of having some meaning assigned.

At the same time, young children may have a much less firm grasp on the meanings of many of their words than adult speakers, and incidental or even irrelevant pragmatic factors may affect their interpretations and responses. Take the case of the Piagetian conservation task where the experimenter ‘checks up’ on the 5- or 6-year-old near-conservers’ answer by asking, for the second time, whether the amount that has just been transferred to a new container or transformed into a different array “is still the same.” Children on the verge of conserving typically change

their initially correct answers from ‘yes’ to ‘no’ at this point. They do so because, pragmatically, asking the same question a second time signals that the initial answer was unsatisfactory (Siegal, 1997).

Another Approach

In another approach to the acquisition of lexical meaning, some researchers have proposed that the task is so complex for young children that they must start out with the help of some *a priori* constraints. These constraints are designed to limit the kinds of meanings children can attribute to new words. What form would these constraints take, and what evidence is there for them?

Among the constraints proposed are whole object – ‘Words pick out whole objects’ – and mutual exclusivity: ‘Each referent is picked by just one word’ (e.g., Markman, 1989). The whole object constraint predicts that young children should assume that any unfamiliar word picks out a whole object and not, for example, a part or property of that object. The mutual exclusivity constraint predicts that young children should assume that an unfamiliar word must pick out something other than whatever has a name that is already known to the child. So this constraint predicts that children will systematically reject second terms they hear apparently applied to an already labeled referent, as well as fail to learn second terms.

The predictions from these and other constraints have been tested in a variety of word-learning experiments where the target referents are objects. In fact, the whole object and mutual exclusivity constraints apply only to words for objects, so they would have infants treat all unfamiliar words as if they designated only objects and never actions, relations, or properties. How do such constraints work, as they conflict with many properties of word meanings? For example, mutual exclusivity would cause children to not learn inclusion relations in taxonomies because they would need to apply two or more terms to the same referent category in learning that an X can be called a *dog*, specifically a subtype called a *spaniel*, and that a dog is also a kind of *animal*. The whole object constraint would cause children to not learn terms for parts and properties. It would militate against children learning any terms for activities or relations.

One could propose that such constraints apply only in the early stages of acquisition, after which they are overridden. However, then one has to specify what leads to their being overridden; in other words, what the necessary and sufficient conditions are for each constraint to be dropped so children can start to learn

words for activities and relations, say, from adult usage, or words for parts and properties, as well as words for objects. Under this view of meaning acquisition, children could start out observing the various constraints and then drop each one at a certain stage in development so as to be able to learn other kinds of meanings up to then blocked by the constraints. In short, children should at first ignore much of what their parents say about words and word meanings and reject second labels whenever they are offered to mark a different perspective, for example. They should also look for words only to pick out objects, mistakenly assigning any that might, in context, seem to be designating categories of actions or relations as words for objects instead.

Is this a realistic picture of development? No, because it calls for selectively ignoring or rejecting a large amount of what adults do with language as they talk about the world to their children, offer them words for objects and events in the locus of joint attention, and provide extensive commentary on parts, properties, motion, and functions associated with specific category members. The constraints approach ignores conditions imposed on conversational exchanges, such as joint attention, and physical and conversational co-presence, and what they contribute to assigning meaning. It also conflicts with adult usage, which offers a range of perspectives on specific objects and events. A piece of fruit can be just that, *fruit*, or it can be *an apple*, *dessert*, or *a snack*, depending on the perspective chosen (Clark, 1997). Yet, these factors must all be taken into account in both designing and interpreting studies of meaning acquisition.

Sources of Meanings

Children draw on conceptual categories already known to them and on information offered in context, both nonlinguistic and linguistic, when they assign a first meaning to new words. Infants build up and organize conceptual categories of the objects, relations, and events they observe months before they try to use words to evoke the relevant categories. As they assign candidate meanings, they rely on these conceptual categories to connect category instances and words as they start in on language (Slobin, 1985). However, because languages differ, children learn, for purposes of talking, to attend automatically to some aspects of events and ignore others; for example, whether an action is complete or not or whether the speaker witnessed an event for herself or simply heard about it. It is therefore important to distinguish between conceptual knowledge about events and the

knowledge speakers draw on when they talk about those events (Slobin, 1996).

Children try to make sense of what the adult wants. This means relying on any potentially useful source of information for interpreting and responding to adult utterances. What children know about the conceptual categories that appear to be at the focus of joint attention seems to provide initial strategies for coping when they do not yet understand all the words. The physical and conversational contexts, with joint attention, identify the relevant 'space' in which to act. This holds just as much for responding to half-grasped requests as for direct offers of unfamiliar words. Children attend to what is physically present, to any familiar words, and to any conceptual preferences. These preferences may include choosing greater amounts over lesser ones, assuming that the first event mentioned is the first to occur, and exchanging one state of affairs for another (Clark, 1997). Such coping strategies may be consistent with the conventional meanings of certain words, so children will appear to understand them when in fact they do not. The match of coping strategies and meanings offers one measure of complexity in acquisition: Matches should be simpler to acquire than cases of mismatch.

Children can draw on what they already know about objects and events, relations, and properties for their world so far. Their current knowledge about both conceptual categories and about their language at each stage offers potential meanings, in context, assignable to unfamiliar words. These preliminary meanings can be refined, added to, and reshaped by adult usage on subsequent occasions. This way, children learn more about the meanings that about each word, the contrasts relevant in particular semantic domains, and the number of terms in a domain that have to be distinguished from one another. To succeed in this effort, children have to identify consistent word uses for specific event-, relation-, and object-types. They have to learn what the conventions are for usage in the speech community where they are growing up (e.g., Eckert, 2003).

Summary

As children learn new words, they rely on what they know so far – the conceptual and linguistic knowledge they have already built up – to assign them some meaning in context. These initial meanings draw equally on their own conceptual categories and on adult patterns of word use within the current conversational exchange. In effect, joint attention, along with what is co-present physically and

conversationally, places pragmatic limits on what the meaning of an unfamiliar word is most likely to be. In addition, adults often provide further information about the referent object or action, linking the word just offered to other words for relevant properties and actions, and thereby situating the new word in relation to terms already known to the child. This framing by adults for new word meanings licenses a variety of inferences by the child about what to keep track of as relevant to each particular word meaning (Clark, 2002). Adults here are the experts and constitute both a source and resource for finding out about unfamiliar word meanings.

See also: Context and Common Ground; Cooperative Principle; Lexical Semantics: Overview; Pragmatic Determinants of What Is Said; Sense and Reference: Philosophical Aspects.

Bibliography

- Bloom P (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Bowerman M (2005). 'Why can't you "open" a nut or "break" a cooked noodle? Learning covert object categories in action word meanings.' In Gershkoff-Stowe L & Rakison D (eds.) *Building object categories in developmental time*. Mahwah, NJ: Lawrence Erlbaum.
- Brown R (1958). *Words and things*. New York: Free Press.
- Carpenter M, Nagell K & Tomasello M (1998). 'Social cognition, joint attention, and communicative competence from 9 to 15 months of age.' *Monographs of the Society for Research in Child Development* 63(176).
- Chouinard M M & Clark E V (2003). 'Adult reformulations of child errors as negative evidence.' *Journal of Child Language* 30, 637–669.
- Clark E V (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Clark E V (1997). 'Conceptual perspective and lexical choice in acquisition.' *Cognition* 64, 1–37.
- Clark E V (2001). 'Grounding and attention in the acquisition of language.' In Andronis M, Ball C, Elston H & Neuvel S (eds.) *Papers from the 37th meeting of the Chicago Linguistic Society*, vol. 1. Chicago: Chicago Linguistic Society. 95–116.
- Clark E V (2002). 'Making use of pragmatic inferences in the acquisition of meaning.' In Beaver D, Kaufmann S, Clark B Z & Casillas L (eds.) *The construction of meaning*. Stanford, CA: CSLI Publications. 45–58.
- Clark E V (2003). *First language acquisition*. Cambridge: Cambridge University Press.
- Clark E V (2004). 'Pragmatics and language acquisition.' In Horn L R & Ward G (eds.) *Handbook of pragmatics*. Oxford: Blackwell. 562–577.
- Clark E V & Grossman J B (1998). 'Pragmatic directions and children's word learning.' *Journal of Child Language* 25, 1–18.
- Clark E V & Wong A D-W (2002). 'Pragmatic directions about language use: words and word meanings.' *Language in Society* 31, 181–212.
- Clark H H (1996). *Using language*. Cambridge: Cambridge University Press.
- Diesendruck G & Markson L (2001). 'Children's avoidance of lexical overlap: a pragmatic account.' *Developmental Psychology* 37, 630–641.
- Eckert P (2003). 'Social variation in America.' *Publication of the American Dialect Society* 88, 99–121.
- Goodman J C, McDonough L & Brown N B (1998). 'The role of semantic context and memory in the acquisition of novel nouns.' *Child Development* 69, 1330–1344.
- Grice H P (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.
- Markman E M (1989). *Categorization and naming in children: problems of induction*. Cambridge, MA: MIT Press.
- Meltzoff A N (1995). 'Understanding the intentions of others: re-enactment of intended acts by eighteen-month-old children.' *Developmental Psychology* 31, 838–850.
- Rogers D (1978). 'Information about word-meaning in the speech of parents to young children.' In Campbell R N & Smith P T (eds.) *Recent advances in the psychology of language*. London: Plenum. 187–198.
- Saylor M M & Sabbagh M A (2004). 'Different kinds of information affect word learning in the preschool years: the case of part-term learning.' *Child Development* 75, 395–408.
- Siegal M (1997). *Knowing children: experiments in conversation and cognition* (2nd edn.). Hove, Sussex: Psychology Press.
- Slobin D I (1985). 'Crosslinguistic evidence for the language-making capacity.' In Slobin D I (ed.) *The crosslinguistic study of language acquisition*, vol. 2. Hillsdale, NJ: Lawrence Erlbaum. 1157–1249.
- Slobin D I (1996). 'From "thought and language" to "thinking for speaking."' In Gumperz J J & Levinson S C (eds.) *Rethinking linguistic relativity*. Cambridge: Cambridge University Press. 70–96.
- Taylor M & Gelman S A (1989). 'Incorporating new words into the lexicon: preliminary evidence for language hierarchies in two-year-old children.' *Child Development* 60, 625–636.
- Tomasello M (1995). 'Joint attention as social cognition.' In Moore C & Dunham P J (eds.) *Joint attention: its origins and role in development*. Hillsdale, NJ: Lawrence Erlbaum. 103–130.
- Tomasello M (2002). 'Perceiving intentions and learning words in the second year of life.' In Bowerman M & Levinson S C (eds.) *Language acquisition and conceptual development*. Cambridge: Cambridge University Press. 132–158.

Meaning: Overview of Philosophical Theories

R M Martin, Dalhousie University, Halifax, NS, Canada

© 2006 Elsevier Ltd. All rights reserved.

The Direct Reference Theory

It is obvious that an important fact about language is that bits of it are systematically related to things in the world. ‘Referential’ theories of meaning hold that the meaning of an expression is a matter, somehow, of this connection.

The most immediately plausible application of this theory is to the meaning of proper names: the name ‘Benedict Spinoza’ is connected to the philosopher, and this fact appears to exhaust the meaning of that name. The other names – ‘Benedictus de Spinoza,’ ‘Baruch Spinoza,’ and ‘Benedict d’Espinosa’ – mean the same because they are connected to the same person.

But even in the case of proper names, problems arise. For example, consider proper names with non-existent references: ‘Santa Claus.’ If the meaning of a proper name is constituted by nothing but its relationship to the bearer of that name, then it follows that ‘Santa Claus’ is meaningless; but this seems wrong, because we know perfectly well what ‘Santa Claus’ means, and we can use it perfectly well, meaningfully. Another example would be the two proper names applied to the planet Venus by the Ancient Greeks, who were unaware that it was the same planet that appeared sometimes in the evening sky, when they called it ‘Hesperus’ and sometimes in the morning sky, when they called it ‘Phosphorus.’ Because these two names in fact refer to one and the same object, we should count them as having exactly the same meaning. It would appear to follow that someone who knew the meaning of both names would recognize that the meaning of one was exactly the same as the meaning of the other, and therefore would be willing to apply them identically. But the Greeks, when seeing Venus in the morning sky, would apply ‘Phosphorus’ to it, but refuse to apply ‘Hesperus.’ Does it follow that they did not fully understand the meanings of those terms? That is an implausible conclusion, since these are terms of the ancient Greek language: how could the most competent speakers of that language fail to understand the meanings of two terms in that language? It looks much more plausible to say that the fact that Hesperus and Phosphorus are identical is not a consequence of the meanings of those words. So meaning is apparently not exhausted by reference. (This example and this argument were originated by Frege.)

But here is a second sort of difficulty for the reference theory. Even if it could do a plausible job of explaining the meaning of proper names, it is not at all clear what it should do with other elements of language. Proper names, after all, make up only a small part of language, and an atypical part, insofar as meaning is concerned, at that; one does not find them in most dictionaries, for example.

Consider whether this sort of approach to meaning might be extended to linguistic items other than proper names. It is a short step from there to more complex, less direct ways of referring, for example, ‘the Amsterdam-born author of the *Ethics*.’ If this definite description gets its meaning by its reference, then since it refers to Spinoza again, it must mean the same as those other names. But a problem immediately arises here, similar to the ‘Hesperus/Phosphorus’ worry. One might understand the meaning of ‘Benedict Spinoza’ perfectly, it seems, without knowing some facts about the philosopher, for example, that he was born in Amsterdam and wrote the *Ethics*; and, as a result, although one understood ‘the Amsterdam-born author of the *Ethics*’ he or she would not know that this referred to Spinoza. A similar problem arises with the ‘Santa Claus’ worry: ‘the present king of France’ is surely meaningful, although it is without reference (Russell’s famous example).

Still other linguistic elements provide greater difficulty for a reference theory. How, for example, is the meaning of the predicate ‘is wise,’ occurring, for example, in ‘Spinoza is wise,’ to be explained in terms of reference? Particular wise objects exist, to be sure – Spinoza for one. But clearly it is not helpful to think that ‘is wise’ here gets its meaning merely by referring to Spinoza again – which would add nothing – or to some other wise person – which seems irrelevant. And what if that sentence were false (but meaningful), and Spinoza were not among the wise things: what would ‘is wise’ refer to then? More likely ‘is wise’ refers to a set of things – the wise individuals (Spinoza, Socrates, Bertrand Russell, etc.). But the sentence clearly tells us more than that Spinoza belongs to the group Spinoza, Socrates, Bertrand Russell, etc. It refers not to that group, it seems, but rather to the property that is necessary for inclusion in that group: wisdom. It is controversial whether we really need to populate our universe with strange objects such as ‘properties’ and ‘universals’; but, in any case, even if they do exist, it’s not clear that ordinary predicates refer to them. For example, ‘wisdom’ may be the name of a particular thing, referred to in the sentence, ‘Wisdom appeals to Francine,’ but it is much less clear that this thing is referred to in the sentence ‘Spinoza is wise.’ A similar

difficulty is posed by common nouns, e.g., ‘philosopher.’ It does not seem that we could explain the meaning of this element in the sentence ‘Spinoza is a philosopher’ by claiming reference to a particular philosopher, the class of philosophers, or philosopher-ness.

Furthermore, reference has nothing to do with grammatical structure, which one would think is an important part of the meaning of any sentence. These two sentences, ‘Jane loves John’ and ‘John loves Jane,’ make the same references (to Jane, John, and loving, perhaps) but they surely mean something very different. A sentence conveys more than a series of references. It does not merely point at Jane, John, and the property of loving; in addition, it makes the claim that Jane loves John (or vice versa).

Meaning as Truth Conditions

Perhaps a more promising way to extend the reference theory for common nouns, predicates, and other linguistic elements is to think of them as functions. Consider the analogy with arithmetic: 5, 13, -9, and so on are the names of numbers (whatever they are), but $x/3 = 9$ is a function from numbers to a ‘truth value.’ With ‘27’ as the argument, its value is TRUE. With ‘16’ as the argument, its value is FALSE. Its meaning consists in the systematic way in which it pairs arguments with truth values. Now consider the systematic way ‘x is wise’ relates arguments to truth values. Substitute the proper name of certain things (any of the wise things) and the value is TRUE. Substitute the proper name of other things and the value is FALSE. The systematic way in which arguments and values are related in this case (it seems) exhausts the meaning of ‘is wise.’ Philosophers have proposed similar ways to deal with other linguistic elements. For example, adverbs might be regarded as functions taking a predicate as ‘argument’ and yielding a predicate as ‘value.’

This amendment in the spirit of the direct reference theory considerably extends its power and explains the function, basically in terms of reference, of various parts of language that do not by themselves refer. Partially because some of the functions in this approach have TRUE and FALSE as values, it was proposed that these truth values be considered the referents of sentences. (This move has seemed implausible to many, however: what are these things called truth values?)

In the 1930s, Alfred Tarski proposed a definition of ‘truth’ that some philosophers thought would be the basis of a good theory of meaning. Tarski’s proposal was that what would constitute a definition of TRUE for a language *L* would be a complete list of

statements giving the truth conditions for each of the sentences in *L*. So one of these statements defining truth-in-English would be ‘*Snow is white*’ is true in English if and only if snow is white. (This may appear ludicrously trivial, because the sentence whose truth conditions are being given, and the reference to the truth condition itself, are in the same language. Of course, if you did not know what made ‘Snow is white’ true, this statement would not tell you. But that is not a problem with Tarski’s view in particular: no statement of a theory would be informative to somebody who didn’t speak the language in which the theory was stated.)

Now, when we know the truth-conditions for a sentence, then do we know its meaning? Once you know, for example, what it takes to make ‘Snow is white’ true, then, it seems, you know what that sentence means. And what it takes, after all, is that snow be white.

Obviously what one learns when one learns the meaning of a language cannot be the truth conditions of each sentence in the language, one at a time, because there are an infinite number of sentences. What is learned must be the meanings of a finite number of elements of sentences, and a finite variety of structures they go into. In the Tarskian view, then, the semantic theory of a language consists of a large but finite number of elements (words, perhaps), together with composition rules for putting them together into sentences and information sufficient for deriving the truth conditions for each of a potentially infinite number of sentences.

One might object that this elaborate theory could not be what people know when they know what their language means. But perhaps providing this is not the purpose of a semantic theory (or a theory of anything). Baseball players are adept at predicting the path of a ball but only rarely are familiar with Newtonian theory of falling bodies.

The idea here is attractive. If you know what sort of world would make a sentence true, then it seems that that is all it would take for you to know what that sentence means. This idea (though not particularly through Tarski’s influence) was the basis of the ‘logical positivist’ theories of meaning and of meaningfulness. Logical positivists enjoyed pointing out that Heidegger’s famous assertion “*Das Nicht nichtet*” (‘The nothing nothings’) was associated with no particular ways the world might be that would make it either true or false, and concluded that this statement, along with many others in metaphysics (e.g., McTaggart’s assertion that time is unreal), were meaningless. But it seemed that baby and bathwater alike were being flushed down the drain. Coupled with a rather narrow and ferocious empiricism, this

criterion ruled out as meaningless a number of assertions that were clearly acceptable. What empirical truth conditions are there **now** for statements about the past, or for assertions about invisible subatomic particles? But this may be a problem more about the logical positivists' narrow empiricism than about their theory of meaning/meaningfulness. More germane here is the problem that many perfectly meaningful sentences have no truth conditions because they're neither true nor false: 'Please pass the salt,' for example.

Sense and Reference

Because of the Hesperus/Phosphorus problem mentioned above, Frege rejected the idea that the meaning of an expression is the thing it refers to. So Frege distinguished the thing to which a symbol referred (in his words, the *Bedeutung*, the 'referent' or 'nominatum') from what he counted as the meaning (the *Sinn*, usually translated as the 'sense') of the symbol, expressed by the symbol. The sense of a symbol, according to Frege, corresponded to a particular way the referent was presented. It might be called the 'way of thinking about' the referent. While his theory separated meaning from reference, nevertheless it can be considered a 'mediated reference theory': senses are ways a reference would be encountered, ways of getting to things from the words that refer to them. But it is the reference of included terms, not their sense, that determines the truth value of the sentence.

Frege's approach led him to a problem with sentences such as these:

- (1) Fred said, "Venus is bright tonight."
- (2) Fred believes he's seeing Venus.

Both sentences include examples of 'opaque context,' a part in the sentence in which substitution of a co-referring term can change the truth value of the sentence. In each of these sentences, substituting 'the second planet from the sun' for 'Venus' may make a true sentence false, or a false one true. In (1), an example of 'direct discourse,' if Fred's very words did not include 'the second planet from the sun,' then that substitution can make a true sentence into a false one. That substitution in (2) may result in a false sentence if Fred believes that Venus is the third planet from the sun. Frege's solution to this problem is to stipulate that in direct discourse – word-for-word quotation – the expression quoted refers to 'itself,' rather than to its usual referent (in this case, Venus). And in belief contexts and some other opaque contexts, expressions refer to their 'senses,' not to their usual referents.

But what, exactly, are these 'senses'? First, it is clear that Frege did not intend them to be the ideas anybody

associates with words. Frege's 'senses' are objective: real facts of the language whose conventions associate senses with its symbols. One may have any sort of association with a bit of language, but the conventions of the language specify only certain of them as meaning-related. (Therefore, Lewis Carroll's Humpty Dumpty does not succeed in meaning 'There's a nice knock-down argument for you' with 'There's glory for you,' even though he insists "When I use a word it means just what I choose it to mean – neither more nor less.") But why not admit that in addition to the public languages there can be 'private' ones with idiosyncratic senses? More will be said about this later.

But second, neither can 'senses' be characteristics of the things referred to: for then, whatever is a sense of 'Hesperus' would be a sense of 'Phosphorus.' Furthermore, there appear to be symbols in a language that have sense but no reference: 'the present king of France' and 'Atlantis' are examples. Reference-less terms pose a problem for Frege. Should he consider them words with sense but no reference? If so, then how can they figure in sentences with a truth-value? (This is similar to the 'Santa Claus' problem.)

A promising approach seems to be to say that the sense of a term 'T' consists of those characteristics judged to be true of things that are called 'T' by competent speakers of the language. But this immediately creates a problem with proper names. If ordinary proper names have senses – associated characteristics that are the way of presenting the individual named, associated conventionally with that name by the language – then there would be corresponding definitional (hence necessary) truths about individuals referred to by proper names. But this is problematic. The sense of a name applying to one individual cannot be the sense of the name of any other individual, because senses are the way terms pick out their designata. So the characteristics associated with a name would have to constitute that individual's 'individual essence' – characteristics uniquely and necessarily true of that individual. But many philosophers have doubted that there are any such characteristics. Even if we can find characteristics that uniquely designate an individual, Kripke (1972) influentially argued that these characteristics are never necessary. Suppose, for example, that 'Aristotle' carries the sense 'Ancient Greek philosopher, student of Plato, teacher of Alexander the Great.' It would follow that this determined the referent of 'Aristotle'; so if historians discovered after all that no student of Plato's ever taught Alexander the Great, then 'Aristotle' would turn out to be a bearer-less proper name, like 'Santa Claus.' But this is not how the proper name would work. Instead,

we would just decide that Aristotle did not teach Alexander after all. Kripke argues that all sentences predicating something of a proper-named individual are contingent, so proper names do not have senses. But, of course, they are meaningful bits of language.

This problem may apply even more broadly than merely to proper names. Consider the meaning of the term ‘water.’ Back before anyone knew the real properties of what made water water – i.e., its chemical constitution – competent speakers applied the term to any colorless, odorless liquid. But they were sometimes wrong, because the characteristics then used to distinguish proper and improper applications of the term, although they happened to pick out water on the whole, were not the genuinely necessary and sufficient conditions for something to be water at all. In those days, then, the real sense of the word was totally unknown and unused by any competent speaker of the language.

Quine argued that Frege’s senses lack what is necessary for well-behaved theoretical objects. We have no idea, for example, of their identity conditions: is the sense of this word the same as the sense of that? More about Quine will be discussed in the final section of this article.

The Idea Theory

The theories discussed so far consider what linguistic elements mean, but other classical theories have concentrated on what people mean by bits of language.

Words, Locke argued, are used as ‘sensible marks’ of ideas; the idea corresponding to a word is its meaning. This has a certain amount of intuitive plausibility, in that non-philosophers think of language as a way of communicating ideas that is successful when it reproduces the speaker’s idea in the hearer’s mind.

The ‘idea theory’ of meaning received its fullest expression in conjunction with British Empiricist epistemology. For the classical empiricists, our ideas were copies of sense-impressions – presumably similar to the sense experiences themselves, except dimmer. These mental representations served us as the elements of thought and provided the meanings for our words.

However, problems with this theory are obvious. For one thing, not every such association is relevant to meaning. For example, the word ‘chocolate’ makes me picture the little shop in Belgium where I bought an astoundingly impressive chocolate bar. But although one might want to say, in a sort of loose way, that that’s what ‘chocolate’ means to me, it doesn’t seem to be a real part of the word’s meaning. Someone else could be completely competent in using that word without any mental pictures of Belgium.

Also, there are some meaningful terms that seem to be associated with no mental imagery, for example, ‘compromise.’ The problem of the meaning of ‘unicorn’ is solvable: images of a horse’s shape and an antelope’s horn can be mentally pasted together to provide a representation; but there are other problems. My image of my cat Tabitha might picture her facing right; but I’m to use this also to identify her as the bearer of that name when she’s facing left; so the mere association of a word with an image is not enough to give that word meaning. There must also be some procedure for using that image. Common nouns (‘dog’) could stand for any and all dogs, whereas the meaning of ‘dog’ was presumably a particular image of a particular dog. More of a theoretical mechanism is needed to explain why this image stands for Fido and Rover but not for Tabitha. And other sorts of words – logical words, prepositions, conjunctions, etc. – raise problems here too: how could sensory experiences be the source of **their** meaning?

A more recent concern about the idea theory arose from the fact that the ideas that gave bits of language their meaning were private entities, whereas the meanings of a public language were presumably public. Clearly I would learn the word ‘cat’ by hearing you use the word in the presence of cats, but not in their absence; but according to the idea theory, I would have learned it correctly when my private image matches yours – something that is impossible for either of us to check. What we can check – identical identifications of objects as cats and non-cats – does not ensure identical private imagery and (according to the idea theory) has nothing directly to do with the meaning we invest ‘cat’ with anyway. Wittgenstein’s ‘private language argument,’ deployed against the idea theory, was considered devastating by many philosophers. Very briefly put, this argument is that the meaning of a public bit of language could not be given by a supposedly necessarily private item, such as a mental representation, because this would make impossible any public check – any real check at all – on whether anyone understood the meaning of a term; and without the possibility of a check, there was no distinction between getting the meaning right and getting it wrong.

Meaning as Use

Wittgenstein’s hugely influential suggestion was that we think instead of sentences as “instruments whose senses are their employments” (1953: 421). Starting in the 1930s and 1940s, philosophers began thinking of the meaning of linguistic items as their potential for particular uses by speakers and attempting to isolate and describe a variety of things that people do with

words: linguistic acts, accomplished through the use of bits of language. One clear theoretical advantage of this approach over earlier ones was its treatment of a much wider variety of linguistic function. Whereas earlier approaches tended to concentrate on information giving, now philosophers added a panoply of other uses: asking questions, giving orders, expressing approval, and so on. This clearly represented a huge improvement on the earlier narrower views, which tried to understand all the elements of language as signs – representations – of something external or internal.

Austin distinguished three kinds of things one does with language: (1) the ‘locutionary act,’ which is the utterance (or writing) of bits of a language; (2) the ‘illocutionary act,’ done by means of the locutionary act, for example, reporting, announcing, predicting, admitting, requesting, ordering, proposing, promising, congratulating, thanking; and (3) the ‘perlocutionary act,’ done by means of the illocutionary act, for example, bringing someone to learn *x*, persuading, frightening, amusing, getting someone to do *x*, embarrassing, boring, inspiring someone. What distinguishes illocutionary acts is that they are accomplished just as soon as the hearer hears and understands what the utterer utters. It is clear that the illocutionary act is the one of these three that is relevant to the meaning of the utterance. The performance of the act of merely uttering a sentence obviously has nothing to do with its meaning. Neither does whatever perlocutionary act is performed: you might bore someone by telling her about your trip to Cleveland or by reciting 75 verses of *The fairie queen*, but the fact that both of these acts may serve to bore the hearer does not show that they are similar in meaning. But similarity in meaning is demonstrated by the fact that two different locutionary acts serve to accomplish the same illocutionary act. For example, ‘Do you have the time?’ and ‘What time is it, please?’ perform the same illocutionary act (a polite request for the time) and are thus similar in meaning.

However, this approach does not deny that what the other theories concentrated on is a significant part of language. In Austin’s classification, one part of many locutionary acts is an act of referring; when I say, “Aristotle was the student of Plato,” I’m probably performing the illocutionary act of informing you, but I’m doing that by means of the locutionary act of referring to Aristotle and Plato. And clearly many linguistic acts include ‘propositional content’: one reports, announces, predicts, admits, requests, orders, proposes, promises, and so on, ‘that *p*,’ so it seems that inside this theory we would need an account of the way any of these linguistic acts correspond to actual or possible states of the world.

A recent influential theory from Donald Davidson responds to these needs by combining, in effect, speech act theory with Tarskian semantics. According to Davidson’s proposal, the list of truth conditions for each assertion in a language provides an account of the language’s semantics – at least, of the propositional content of sentences in the language: what a statement, prediction, assertion, question, etc., states, predicts, asserts, asks, etc. This explains what is shared by ‘The light is turned off,’ ‘Turn off that light,’ ‘Is the light turned off?’ and so on. But secondly, according to Davidson, a theory of meaning needs a ‘mood indicator’ – an element of the sentence that indicates the use of that sentence – e.g., as a statement, request, or question.

Quine’s Skepticism

Quine’s skepticism about meanings was among the most important 20th-century developments in the philosophy of language. Here is one way of approaching his position. Imagine a linguist trying to translate a tribe’s language. Suppose that the tribesmen said “*gavagai*” whenever a rabbit ran past. Does ‘*gavagai*’ mean ‘lo, a rabbit!’? The evidence might as well be taken to show that ‘*gavagai*’ asserts the presence of an undetached rabbit part, a temporal slice of a rabbit, or any one of a number of other alternatives, including even ‘Am I ever hungry!’ or ‘There goes a hedgehog!’ (if the tribesmen were interested in misleading you). Quine called this the ‘indeterminacy of translation.’ But, we might object, couldn’t further observation and experimentation decide which of these alternatives is the right interpretation? No, argued Quine, there are always alternatives consistent with any amount of observation. But, we might reply (and this is a more basic objection), what that shows is that a linguist never has absolutely perfect evidence for a unique translation. This is by now a familiar (Quinian) point about theory: theory is always to some extent undetermined by observation. In any science, one can dream up alternative theories to the preferred theory that are equally consistent with all the evidence to date. But in those cases, isn’t there a right answer – a real fact of the matter – which, unfortunately, we may never be in a **perfect** position to determine, because our evidence must always be equivocal to some extent? At least in the case of meaning, argued Quine, the answer is no, because for Quine, linguistic behavior is all there is to language, so there are no hidden facts about meaning to discover, with linguistic behavior as evidence. So meanings are not given by ideas in the head, Fregean senses, or anything else external to this behavior.

A similar sort of position was argued for more recently by Kripke (1982). Imagine that

someone – Fred – has used a word ‘W’ to refer to various things, A, B, and C. Now he encounters D: is that referred to by ‘W’? One wants to say: if D is like A, B, and C, then Fred should go on in the same way and apply ‘W’ to D. But Kripke argues that there is no fact about Fred’s intentions or past behavior – no fact about what he means by ‘W’ – that would make it correct or incorrect for him to apply ‘W’ to D. Neither is there, in the external world, a real sufficient (or insufficient) similarity of D to A, B, and C that make it correct (or incorrect). The only thing that would make that application correct or incorrect is what a community of speakers using that word would happen to agree on.

But does anti-realism about meaning really follow from these considerations? Suppose that Quine and Kripke are right, and all there is to language is social behavior. But maybe this does not imply that meanings are unreal. When we consider an action as social behavior, after all, we do not think of it (as Quine, in effect, did) merely as bodily movements. There are facts about the social significance of the behavior, above and beyond these movements, that give the movement its social significance. Perhaps it is these facts that would make one linguistic theory rather than another correct – that determine the meaning of the noises made by Quine’s natives, and whether Fred is following the linguistic rule when he applies ‘W’ to D. Language is a tool we know how to use, and the real meaning of our utterances is what we know when we know how to use that tool.

See also: Descriptions, Definite and Indefinite: Philosophical Aspects; Direct Reference; Empiricism; Empty Names; Expression Meaning versus Utterance/Speaker Meaning; Ideational Theories of Meaning; Indeterminacy, Semantic; Intention and Semantics; Mood, Clause Types, and Illocutionary Force; Nominalism; Private Language Argument; Proper Names: Philosophical Aspects; Radical Interpretation, Translation and Interpretationalism; Realism and

Antirealism; Reference: Philosophical Theories; Rigid Designation; Sense and Reference: Philosophical Aspects; Speech Acts; Thought and Language: Philosophical Aspects; Truth Conditional Semantics and Meaning; Use Theories of Meaning; Use versus Mention.

Bibliography

- Alston W P (1964). *Philosophy of language*. Englewood Cliffs, NJ: Prentice-Hall.
- Austin J L (1962). *How to do things with words*. Urmson J O & Sbisá M (eds.). Cambridge, MA: Harvard University Press.
- Blackburn S (1984). *Spreading the word: groundings in the philosophy of language*. Oxford: Clarendon Press.
- Davidson D (1967). ‘Truth and meaning.’ *Synthese* 17, 304–323.
- Frege G (1892). ‘On sense and reference.’ In Geach P & Black M (eds.) *Translations from the philosophical writings of Gottlob Frege*. Oxford: Basil Blackwell.
- Grice H P (1957). ‘Meaning.’ *Philosophical Review* 66, 377–388.
- Kripke S (1972). ‘Naming and necessity.’ In Davidson D & Harmon G (eds.) *Semantics of natural language*. Dordrecht: Reidel.
- Kripke S (1982). *On rules and private language*. Cambridge, MA: Harvard University Press.
- Martin R M (1987). *The meaning of language*. Cambridge, MA: The MIT Press.
- Mill J S (1872). *A system of logic, book I* (8th edn.). London: Longmans.
- Quine W V O (1951). ‘Two dogmas of empiricism.’ In *From a logical point of view*. Cambridge, MA: Harvard University Press.
- Quine W V O (1960). *Word and object*. Cambridge, MA: The MIT Press.
- Russell B (1905). ‘On denoting.’ *Mind* 14, 479–493.
- Searle J R (1969). *Speech acts*. Cambridge: Cambridge University Press.
- Stainton R J (1996). *Philosophical perspectives on language*. Peterborough, ON: Broadview Press.
- Wittgenstein L (1953). *Philosophical investigations*. Anscombe G E M (trans.). Oxford: Basil Blackwell.

Meaning: Procedural and Conceptual

D Blakemore, University of Salford, Greater Manchester, UK

© 2006 Elsevier Ltd. All rights reserved.

The distinction between procedural and conceptual meaning resulted from the attempt to reanalyze Grice’s (1989) notion of conventional implicature in relevance theoretic terms, or, more generally, from

the attempt to provide a cognitive reanalysis of the distinction drawn between truth conditional and non-truth conditional meaning (see Grice, 1989; see **Implicature**). However, soon after its introduction (Blakemore, 1987), it was shown that the relevance theoretic distinction was not coextensive with the distinction between truth conditional and non-truth conditional meaning, and thus represented a departure from the approach to semantics that

underlies Grice's notion of conventional implicature. While there is a range of phenomena that can be analyzed in terms of the conceptual–procedural distinction, the emphasis in pragmatics has mostly centered on its application to the analysis of so-called 'discourse markers'. It has to be said, however, that the term *procedural* is not always used in the same way, and that the distinction between conceptual and procedural encoding is not always drawn within a relevance theoretic framework. In argumentation theory, any expression which has an argumentative function is said to encode procedural meaning (Moeschler, 1989). According to this criterion, *because* is procedural; however, it is not procedural according to the relevance theoretic definition outlined below. Fraser (1996, 1998) has used the term *procedural* to describe the meaning of any expression with an indicating function, where *indicating* is intended in its speech act theoretic sense, in which it contrasts with *saying* or *describing*. Because the relevance theoretic distinction is not coextensive with the traditional distinction between truth conditional and nontruth conditional meaning, it is not surprising that there are expressions which are procedural according to Fraser but conceptual according to the relevance theoretic definition – e.g., *as a result*. This article will focus on the relevance theoretic distinction, showing, on the one hand, how it derives from a relevance theoretic view of linguistic semantics, and on the other, how the principle of relevance provides an explanation for the fact that languages have developed means for encoding procedures (see **Relevance Theory**). It will then turn to the way in which procedural encoding has been applied in the analysis of discourse markers or connectives, and the questions this type of analysis raises.

Relevance Theoretic Semantics

Grice's (1989) notion of conventional implicature was a means of maintaining a definition of *what is said* in which linguistic meaning coincides with truth conditional content. While he accepts that the suggestion carried by *therefore* in (1) is linguistically encoded, he does not wish to allow that in his "favoured sense of 'say,' one who utters (1) would have *said* that Bill's being courageous follows from his being a philosopher" (1989: 21):

- (1) Bill is a philosopher and he is, therefore, brave.

Grice's solution to the problem posed by nontruth conditional linguistic meaning was to modify his definition of *what is said* so that it applied only to the performances of so-called *central* speech acts, and not to the performances of *noncentral* speech acts indicated by expressions such as *therefore* (see **Speech Acts**).

However, as Sperber and Wilson (1995) and Carston (2002) have shown, the assumption that linguistic meaning coincides with the truth conditional content of the utterance cannot be justified, because the propositional content of utterances is underdetermined by their linguistic properties. While the thoughts communicated by utterances may have truth conditions, it cannot be said that linguistic form encodes thoughts. Linguistic meaning is an input to the pragmatically constrained inferential processes that use contextual information to deliver thoughts. The question for linguistic semantics, in this view, is not about the relationship between language and the world but about the relationship between linguistic form and the inferential processes that deliver thoughts.

The distinction between conceptual and procedural encoding derives from the argument that if utterance interpretation involves the construction and inferential manipulation of propositional representations, then it is reasonable to expect two answers to this question. On the one hand, linguistic form can encode the constituents of the conceptual representations that take part in inferential computations, and on the other, it can encode information that makes particular kinds of computations salient. Consider (2a) and (2b), where (2b) could be interpreted as a contextual implication derived in an inference that has (2a) as a premise, or as a premise that has (2a) as a conclusion.

(2a) Tom will be late.

(2b) He is coming from London.

The claim that linguistic meaning can encode procedures is the claim that there are linguistic expressions e.g., – *so* or *after all* in this case – which guide the hearer to the inferential procedure that yields the intended interpretation.

Why Languages Develop Procedural Encoding

It has been argued that the fact that languages have developed expressions which constrain inferential procedures can be explained within relevance theory in terms of the communicative principle of relevance (Sperber and Wilson, 1995). According to this principle, a hearer who recognizes that a speaker has deliberately communicated with her is entitled to assume not just that the utterance is relevant enough (in terms of effects and effort) to be worth her attention, but that it is the most relevant utterance the speaker could have produced, given her interests and abilities. Because the degree of relevance is affected by the amount of processing effort required for the derivation of the intended cognitive effects, this means that the use of an expression that encodes a procedure for

identifying the intended interpreted would be consistent with the speaker's aim of identifying relevance for a minimum processing cost.

Traugott and Dasher (2002) have argued that languages tend to develop procedural constraints on interpretation out of their existing conceptual resources. However, although they agree that there is a distinction between procedural and conceptual (or, as they call it, 'contentful') meaning, their account of how meaning change takes place is based on the assumptions of cognitive semantics and neo-Gricean pragmatics rather than on relevance theoretic pragmatics, and it is not clear that the distinction is drawn in the same way. Moreover, their account assumes that meaning, changes results from the conventionalization of pragmatic inferences, where this is unpacked in (Gricean) terms of the conventionalization of the relation between a linguistic form and the proposition or propositions derived by inference. A relevance theoretic account would approach this process in terms of the conventionalization of an inferential routine or process.

If procedural meaning develops from existing conceptual resources, then it would not be surprising for an expression to encode both a concept and a procedure. Although it has been argued that this possibility has been ruled out in relevance theory, it is consistent with relevance theoretic assumptions, and there have been relevance theoretic analyses (e.g., Blakemore's 1987 analysis of *but*, Nicolle's 1997 and 1998 account of *be going to*, and Wilson's 2004 analysis of expressions such as *few* and *several*) where it is argued that a single form may encode two types of meaning.

The Conceptual–Procedural Distinction and Conventional Implicature

While the notion of procedural encoding was developed as a means of analyzing expressions which encode constraints on the recovery of implicit content – e.g., *therefore*, *so*, *after all* – subsequent investigation suggested that it could be extended to expressions that encode constraints on the recovery of explicit content. Some of these – mood indicators, attitudinal particles – are analyzed as encoding constraints on the inferential processes involved in the recovery of higher-level explicatures. For example, Wilson and Sperber (1993) suggest that the use of *huh* in (3) encourages the hearer to construct the higher-level explicature in (4) (see also Wilson and Sperber, 1988; Clark, 1993):

(3) Peter's a genius, huh!

(4) The speaker of (3) doesn't think that Peter is a genius.

In this case, the equation between procedural meaning and nontruth conditional meaning is maintained, because higher-level explicatures are not regarded as contributing to truth conditional content. However, it has also been suggested that there are expressions, notably pronouns, which should be analyzed as constraints on the proposition expressed, and thus contribute to truth conditional content.

At the same time, it has been argued (Wilson and Sperber, 1993; Ifantidou-Trouki, 1993; Blakemore, 1990, 1996) that expressions such as sentence adverbials or parentheticals which, although they do not contribute to truth conditional content, must be analyzed as encoding concepts. This means that the procedural–conceptual distinction cannot be coextensive with the distinction that underlies Grice's notion of conventional implicature.

If Carston's (2002) conception of linguistic semantics is right, then the fundamental distinction must be the conceptual–procedural distinction rather than the distinction between truth conditional and non-truth conditional meaning. This means that linguistic semantics must include a means of distinguishing conceptual from procedural meaning. Within relevance theory, attention has been drawn to properties that distinguish expressions that encode procedures from those that encode concepts. First, in contrast with concepts, procedures cannot be brought to consciousness (Wilson and Sperber, 1993: 16). This explains why even native speakers find it difficult to judge whether expressions which encode procedures – e.g., *but* and *however* in English or *dakara* and *sorede* in Japanese – are synonymous without testing their intersubstitutability in all contexts. In contrast, even when the definition of a concept proves controversial, speakers can say whether two expressions encode the same concept without having to test whether they can be substituted for each other in every context. If it is difficult for native speakers to make synonymy judgments, then it is not surprising that the translation of expressions that encode procedures is notoriously difficult, particularly since languages do not necessarily conventionalize the same inferential routines. However, the translation of procedural meaning has yet to be investigated systematically.

Second, while expressions that encode concepts can be semantically complex, expressions that encode procedures cannot combine with other expressions to produce semantically complex expressions. Compare (5) with (6):

(5) In total, absolute confidence, she has been promoted.

(6) Sue likes red wine. *Totally however, Mary drinks beer.

Rouchota (1998) has shown that while expressions that have been analyzed as encoding procedures can combine in some way, they do not combine to form complex discourse markers. Compare (7) with (5):

- (7) Sue fell asleep during the lecture. But after all, she had heard it all before.

Procedural Analyses of Discourse Markers

The notion of procedural encoding has been applied to the analysis of a range of nontruth conditional discourse markers in a variety of languages (e.g., Blass, 1990; Itani, 1993; Unger, 1996; Matsui, 2002; Iten, 2000; Blakemore, 2002). However, this work has suggested that the original notion is neither sufficiently fine-grained to capture the differences between closely related but nonintersubstitutable discourse markers (e.g., *but* and *however* or *dakara* and *sorede*), nor sufficiently broad to capture all the ways in which linguistic form can encode information about the inferential computations involved in the interpretation of utterances. As it is defined by Blakemore (1987), procedural encoding is linked to the three cognitive effects defined within relevance theory: contextual implication, strengthening, and elimination. In order to account for the differences between expressions such as *but* and *however*, it has been argued that the notion of procedural encoding must be broadened to include the activation of particular types of contexts. Moreover, it has been suggested that the meanings of some discourse markers (e.g., *well*) may not necessarily be linked to the activation of cognitive effects at all (see Blakemore, 2002).

Future Directions

Procedural encoding has also played a role in the analysis of the role of intonation in interpretation (see Fretheim, 2002; House, 2004). If, as Gussenhoven (2002) has argued, aspects of intonation have become grammaticalized so that certain pitch contours encode arbitrary meanings, then it is plausible that these meanings should be analyzed in terms of instructions for interpretation. However, Wharton's (2003a, 2000b) work on natural codes suggests it may also be possible to generalize the notion of procedural meaning to accommodate phenomena that are on the borderline of language (e.g., interjections) as well as natural or paralinguistic aspects of intonation. Research in this area has yet to be developed, but it seems clear that the scope of procedural encoding extends beyond the analysis of the nontruth conditional discourse markers.

See also: Implicature; Relevance Theory; Speech Acts.

Bibliography

- Blakemore D (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore D (1990). 'Performatives and parentheticals.' *Proceedings of the Aristotelian Society* 91.3, 197–213.
- Blakemore D (1996). 'Are apposition markers discourse markers?' *Journal of Linguistics* 32, 325–347.
- Blakemore D (2002). *Relevance and linguistic meaning: the semantics and pragmatics of discourse markers*. Cambridge, UK: Cambridge University Press.
- Blass R (1990). *Relevance relations in discourse: a study with special reference to Sissala*. Cambridge: Cambridge University Press.
- Carston R (2002). *Thoughts and utterances: the pragmatics of explicit communication*. Oxford: Blackwell.
- Clark B (1993). 'Relevance and pseudo-imperatives.' *Linguistics and philosophy* 16, 79–121.
- Fraser B (1996). 'Pragmatic markers.' *Pragmatics* 6, 167–190.
- Fraser B (1998). 'Contrastive markers in English.' In Jucker A & Ziv Y (eds.) *Discourse markers: descriptions and theories*. Amsterdam: John Benjamins. 301–326.
- Fretheim T (2002). 'Intonation as a constraint on inferential processing.' In *Proceedings of Speech Prosody 2002*, University of Aix-en-Provence.
- Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Gussenhoven C (2002). 'Intonation and interpretation: phonetics and phonology.' In *Proceedings of Speech Prosody 2002*, University of Aix-en-Provence, 47–57.
- House J (2004). 'Constructing a context with intonation.' Paper read at the 6th NWCL International Conference, *Prosody and Pragmatics*, Preston, UK.
- Ifantidou-Trouki E (1993). 'Sentential adverbs and relevance.' *Lingua* 90, 69–90.
- Itani R (1993). 'The Japanese particle *ka*: a relevance theoretic approach.' *Lingua* 90, 129–147.
- Iten C (2000). 'Non-truth conditional meaning, relevance and concessives.' Ph.D. thesis, University of London.
- Matsui T (2002). 'Semantics and pragmatics of a Japanese discourse marker: *dakara*.' *Journal of Pragmatics* 34.7, 867–889.
- Moeschler J (1989). *Argumentation, relevance and discourse*. *Argumentation* (3.3). Paris: Hermès.
- Nicolle S (1997). 'A relevance theoretic account of *be going to*.' *Journal of Linguistics* 33, 355–377.
- Nicolle S (1998). 'A relevance theoretic perspective on grammaticalization.' *Cognitive Linguistics* 9, 1–35.
- Rouchota V (1998). 'Procedural meaning and parenthetical discourse markers.' In Jucker A & Ziv Y (eds.) *Discourse markers: descriptions and theories*. Amsterdam: John Benjamins. 301–326.
- Sperber D & Wilson D (1995). *Relevance: communication and cognition*. Oxford: Blackwell.
- Traugott E & Dasher R (2002). *Regularity in semantic change*. Cambridge: Cambridge University Press.

- Unger C (1996). 'The scope of discourse connectives: implications for discourse organization.' *Journal of Linguistics* 32, 403–438.
- Wharton T (2003a). 'Interjections, language and the 'showing-saying' continuum.' *Pragmatics and Cognition* 11.1, 39–91.
- Wharton T (2003b). 'Natural pragmatics and natural codes.' *Mind and Language* 18.4, 447–477.
- Wilson D (2004). 'Relevance and argumentation theory.' Paper delivered at *Pragmatic Interfaces*, Geneva.
- Wilson D & Sperber D (1988). 'Mood and the analysis of non-declarative sentences.' In Dancy J, Moravcsik J & Taylor C (eds.) *Human agency: language, duty and value*. Stanford, CA: Stanford University Press.
- Wilson D & Sperber D (1993). 'Linguistic form and relevance.' *Lingua* 90, 1–25.

Mentalese

F Egan, Rutgers University, New Brunswick, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

The Basic Hypothesis

Some theorists of mind have claimed that thought takes place in a language-like medium. They have called this language 'Mentalese.' Mentalese has a syntax, a semantics, and a morphology, though discovering these properties of the language of thought will likely require extensive empirical investigation of the mind. Obviously, mentalese does not have a phonology. It is therefore more like written public language than overt speech. And whereas public languages require a pragmatics – a theory of how the language is used by speakers – mentalese, like the machine languages of computers, does not call for one.

Gilbert Harman (1973) offered the following argument for the existence of mentalese: logical relations hold among mental states, and these relations are essential to their role in psychological prediction and explanation. If the belief that *snow is white and grass is green* is true, then the belief that *snow is white* is true. In general, if the belief that *p & q* is true, then the belief that *p* is true. Generalizations of this sort presuppose that beliefs have sentential structure. Some beliefs are conjunctions, others disjunctions, and so on. Beliefs (as well as desires, fears, and the other propositional attitudes) are part of a language-like system.

Harman's argument fails to establish that mental states themselves have logical or sentential structure. The argument trades on the fact that belief ascriptions have sentential structure. We ascribe certain beliefs to subjects using sentences that are conjunctive or disjunctive, but it does not follow that the mental states so ascribed are themselves conjunctions or disjunctions, or that the relations that hold among these mental states are of the sort that hold among sentences (or propositions), that is, that they are logical relations. To assume that they are is just to assume

what is at issue – that thoughts have a language-like structure. In general, one must guard against attributing to thoughts themselves properties of the representational scheme that we use to talk about them.

The hypothesis that thought occurs in a language-like medium is understood as the claim that not only beliefs but also desires and the other propositional attitudes are properly construed as relations to sentences in the inner language. To believe that the conflict in the Middle East will not be resolved is to bear a relation to an inner sentence token that means *the conflict in the Middle East will not be resolved*. To fear that the conflict in the Middle East will not be resolved is to bear a different relation to an internal sentence of the same type. The difference between the believing-relation and the fear-relation is construed as a difference in the processing that the sentence token undergoes in the brain, in other words, as a difference in its functional role. The belief is likely to cause, in certain circumstances, sincere assertions of a public language sentence meaning *the conflict in the Middle East will not be resolved*. The fear is more likely to give rise to appropriate emotional states.

What Is Mentalese Like?

The Thinker's Public Language, or a Proprietary Inner Code?

Some theorists have supposed that mentalese is just the thinker's own public language. English speakers think in English, Chinese speakers in Chinese. One might be inclined to this view by reflecting on the fact that many thoughts do not seem possible until the thinker has acquired a public language. Consider, for example, the thought that *the leech of the genoa is curled*. A subject cannot genuinely think this thought until she has acquired the concepts *leech* and *genoa*. Moreover, acquiring such concepts seems to require learning the appropriate public language terms for them, or more basic public language terms in which

they can be defined. If mentalese is just the thinker's public language, then investigation of its properties is relatively straightforward. The syntax of an English speaker's mentalese is just English syntax; its semantics is just English semantics.

Jerry Fodor, in his groundbreaking 1975 book, *The language of thought*, argued that thought takes place in a proprietary inner code; moreover, this inner language has the expressive power of any public language a thinker is capable of learning. According to Fodor, the process of language learning involves hypothesis formation and testing; in particular, the hypotheses have the form of 'truth rules' for the application of the public language terms. To learn the English term 'genoa' is to learn a rule of the form "x is a genoa' is true if and only if x is P," where P is a predicate in the proprietary inner language that is coextensive with the English predicate 'genoa.' To learn a public language is to acquire a translation manual that pairs terms in the language of thought with their public language equivalents. On pain of regress, terms in the language of thought are not themselves learned. A consequence of Fodor's view is that the concept *genoa* – and *electron*, *carburetor*, and any other concept a thinker can possess – is either innate or definable in terms of other concepts that are themselves innate. Fodor argues (Fodor, 1981) that few concepts are so definable; hence the vast majority of a thinker's concepts are innate.

Needless to say, many have found Fodor's extreme nativism unpalatable. His argument depends upon construing (public) language learning as hypothesis formation and confirmation, a process that requires an internal medium of representation, a language of thought where the hypotheses are couched. But there is nothing inevitable about explicit hypothesis formation and testing models of learning. If public language predicates were learned as a result of a causal process that is not construed as linguistic or inferential – a process sometimes known as 'triggering' – then the argument's nativist conclusion would not follow.

Whatever their views on concept-nativism, most proponents of the view that thought takes place in a linguistic medium have followed Fodor in claiming that the language of thought is an inner neural code, distinct from any public language. Accordingly, we will hereinafter construe the 'language of thought hypothesis' (LOT) as the view that thought takes place in a proprietary inner code.

Psycho-Syntax and Psycho-Semantics

If the language of thought is indeed an inner neural code, then discovering its syntax and semantics will require extensive empirical investigation of the mind. Characterizing the syntax of mentalese will involve a

specification of a finite set of primitive objects (words), and a grammar or set of formation rules that describe the ways in which complex syntactic objects (sentences) may be built out of the primitives. The individuation of syntactic types will be functionally based, adverting to the causal roles of these objects in the subject's cognitive life.

The so-called 'mental logic' research program that attempts to uncover the formal rules of inference underlying human deductive reasoning presupposes the existence of an innate syntax of thought and proposes to empirically investigate it. (See the papers in Braine and O'Brien, 1998.) Various candidate inference schemas have been offered, but no proposal is sufficiently detailed to generate empirically testable predictions regarding the underlying syntax.

A full theory of mentalese also requires a 'psycho-semantics' – an account of how internal sentences acquire their meaning. In virtue of what fact does a particular mental sentence mean *snow is white* rather than $2+2=4$? The meanings of public language sentences are fixed by public agreement, or derive in some way from the mental states of the users of these sentences, but, on pain of circularity, the sentences of mentalese must acquire their meanings in some other way. Since the 1980s, there has been a proliferation of theories, mostly by philosophers, purporting to explain how mental representation is possible. Not all of these accounts presuppose LOT, but most, if not all, are compatible with it. Typically, such theories attempt to explain the semantic properties of thought while respecting a naturalistic constraint – they attempt to specify sufficient conditions for a mental state's meaning, what it does in nonsemantic and nonintentional terms. (See Stich and Warfield, 1994 and *Representation in Language and Mind*.)

Further Arguments for LOT

Theories of Mental Processing Are Committed to LOT

Fodor (1975) reasoned that the only plausible models of mental processing are computational models, and that these require a medium of computation, that is, an inner system of representation. This argument has now been undermined by the existence of connectionist computational models. Connectionist machines are capable of performing cognitive tasks, but they lack fixed symbols over which computational operations are defined. Connectionist processes are not naturally interpretable as manipulations of internal sentences or data structures. If the mind is best characterized as a connectionist machine, or as an aggregate of such machines without an overarching executive control, then the LOT hypothesis is false.

LOT Explains Some Pervasive Features of Thought

Fodor (1987) argues that LOT provides the best, indeed, the only explanation of two pervasive features of thought. Thought is **productive**: we can think arbitrarily many thoughts. It is also **systematic**; cognitive capacities are systematically related. If a subject can think the thought *John loves Mary*, then he can think the thought *Mary loves John*. The explanation for the productivity and systematicity of thought is that thoughts have a language-like structure. We can think arbitrarily many thoughts for the same reason that we can utter arbitrarily many sentences. Thoughts, like sentences, are composed of a finite base of elements put together in regular ways, according to the rules of a grammar. The systematicity of thought is also explained by LOT: systematically related thoughts contain the same basic elements but are arranged differently.

Whether the argument is successful depends on two issues: (1) whether productivity and systematicity are indeed pervasive features of thought; and (2) if they are, whether they can be accounted for without positing a language of thought.

Thoughts are assumed to be productive because they are represented, described, and attributed by public language sentences, a system which is itself productive. However, as noted above, one must be careful not to attribute to thoughts themselves properties of the representational scheme that we use to talk about them. It would be a mistake to think that temperature is infinite because the scheme used to measure it, the natural numbers, is infinite. If thoughts are understood as internal states of subjects that are, typically, effects of external conditions and causes of behavior, then it is not obvious that there are arbitrarily many of them. The size of the set of possible belief-states of human thinkers, like the size of the set of possible temperatures of objects, is a matter to be settled by empirical investigation.

Turning to systematicity, the argument falls short of establishing the existence of mentalese. In the first

place, it is not clear how pervasive systematicity really is. It is not generally true that if a thinker can entertain a proposition of the form aRb , then she can entertain bRa . One can think the thought *the boy parsed the sentence*, but not *the sentence parsed the boy*. Moreover, it is a matter of some dispute within the cognitive science community whether connectionist cognitive models, which do not posit a language of thought, might be capable of explaining the systematic relations that do hold among thoughts. (See MacDonald and MacDonald, 1995 for the classic papers on this issue, and Matthews, 1997 for further discussion).

See also: Representation in Language and Mind.

Bibliography

- Braine M & O'Brien D (eds.) (1998). *Mental logic*. Hillsdale, NJ: Laurence Erlbaum Associates.
- Egan F (1991). 'Propositional attitudes and the language of thought.' *Canadian Journal of Philosophy* 21, 379–388.
- Field H (1978). 'Mental representation.' In Block N (ed.) *Readings in the philosophy of psychology*, vol. 2. Cambridge, MA: Harvard University Press. 78–114.
- Fodor J A (1975). *The language of thought*. New York: Thomas Y. Crowell.
- Fodor J A (1981). 'The present status of the innateness controversy.' *RePresentations: Philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press. 257–316.
- Fodor J A (1987). 'Why there still has to be a language of thought.' *Psychosemantics*. Cambridge, MA: MIT Press. 136–154.
- Harman G (1973). *Thought*. Princeton, NJ: Princeton University Press.
- MacDonald C & MacDonald G (1995). *Connectionism: debates on psychological explanation*. Oxford: Blackwell.
- Matthews R J (1997). 'Can connectionists explain systematicity?' *Mind and Language* 12, 154–177.
- Stich S P & Warfield T A (1994). *Mental representation: a reader*. Oxford: Blackwell.

Metalanguage versus Object Language

K Allan, Monash University, Victoria, Australia

© 2006 Elsevier Ltd. All rights reserved.

Human languages are the objects studied in linguistics. For that reason, the language under investigation is known as the object language. The language that a linguist uses to describe and analyze the object

language is called the metalanguage. The basic requirement for a metalanguage is that it satisfactorily communicates the form, structure, and meaning of item e_{OL} – that is, any expression in the object language, whether it is a word, a phrase, or a sentence – in terms of an expression ' e_M ' in the metalanguage. This short article focuses on semantic metalanguages.

Ideally, a metalanguage is a formal language, as in (2); (1) is informal.

- (1) *dog* means ‘canine animal’
 (2) $\forall x[D(x) \leftrightarrow \lambda[A(y) \wedge C(y)](x)]$

Dog is e_{OL} and ‘canine animal’ is ‘ e_M ’; note the typographical differentiation. Example (2) makes recourse to natural language vocabulary because the *D*, *A*, and *C* symbolize *dog*, *animal*, and *canine*; (2) can be read ‘Everything that is a dog is a member of the set of individuals that are both animal and canine’. Strictly, a formal language has a fully defined vocabulary and syntax. Ideally, the vocabulary is a specified set of symbols whose forms and correlated meanings are fully defined, all possible combinations of vocabulary items in the metalanguage are generated from fully specified syntactic axioms and rules of syntax, and the meanings of syntactically well-formed structures are fully specified by semantic axioms and rules for the metalanguage.

A metalanguage is, in effect, a translation of the object language (Carnap, 1937: 228); so, defining a formal metalanguage for natural language semantics requires that it have the same expressive power as a natural language. In order for the metalanguage to be understood and used by human beings, it must be communicable and, hence, translate into a natural language. If you understand neither Polish nor Swahili, there is little point in my using Swahili as a metalanguage for the semantic analysis of Polish (or vice versa); for example, my saying that *to jest pies* means ‘ni mbwa’ will not help you at all. Using English as a metalanguage, I would say *to jest pies* means ‘it’s a dog’; using English as a metalanguage, I would say *to jest pies* means ‘ni mbwa’ (in Swahili), which means ‘it’s a dog.’

The advantages of a formal metalanguage are the explicit definition of primitives and standards of rigor and exactitude that tend to be ignored when using an informal metalanguage. The proper formalization of the metalanguage permits proofs of particular conclusions about semantic structure and so prevents mistaken conclusions deriving from faulty assumptions and/or inference procedures. However, none of these advantages of a formal system is necessarily unobtainable using an informal system such as a natural language metalanguage for semantics. Given two metalanguages with the same descriptive and explanatory capacities, the only way to choose between them is to be guided by gut feeling: favor the one you are happier with. A metalanguage is the product of an analyst’s mind; the analyst not being God, every metalanguage is limited by the beliefs, perspectives, and purposes of its creator.

See also: Formal Semantics.

Bibliography

- Carnap R (1937). *Logical syntax of language*. London: Routledge and Kegan Paul.
 Gamut L T F (1991a). *Language logic and meaning 1: Introduction to logic*. Chicago, IL: Chicago University Press.
 Gamut L T F (1991b). *Language logic and meaning 2: Intensional logic and logical grammar*. Chicago, IL: Chicago University Press.
 Hofstadter D R (1980). *Gödel, Escher, Bach: an eternal golden braid*. Harmondsworth, UK: Penguin.
 Lyons J (1995). *Linguistic semantics: an introduction*. Cambridge, UK: Cambridge University Press.

Metaphor: Philosophical Theories

M Arseneault, University of Wisconsin-Madison, Madison, WI, USA

© 2006 Elsevier Ltd. All rights reserved.

Metaphor and Philosophy

Rather than simply interpreting particular metaphorical expressions, philosophers of language investigate questions such as whether metaphors mean what they do in virtue of semantic content, or whether pragmatic features of expression use and context determine the meaning of metaphors. Answering such questions requires addressing fundamental issues in language

and communication such as issues about the limits of literal meaning and the semantic-pragmatic distinction.

Defining Metaphor

Metaphor theorists disagree about which class of expressions constitutes the proper object of an analysis of metaphor. While most metaphor theorists favor live metaphors (metaphors that invite a multitude of interpretations), others such as George Lakoff and Mark Johnson (1980) focus on dead metaphors (metaphors with a relatively fixed meaning) and their role in structuring cognition. At the same time, some metaphor

theorists adopt broader definitions while others work with narrower definitions of metaphor.

Metaphor, in its broadest sense (metaphor₁), includes most if not all figurative language such that the principle contrast is between metaphorical and literal language. Narrower definitions take metaphor to be only one among many other non-literal tropes. A second type of metaphor (metaphor₂), distinct from other non-literal tropes such as irony, metonymy, and synecdoche, is what makes us think of one thing *as* another. Since the very possibility of giving a unified account of metaphor₁ is remote, articulating an account of metaphor₂ is a more appropriate goal. However, a demarcation problem remains a challenge for any treatment of metaphor₂: it must be worked out what it means to say that metaphor₂ makes us think of one thing as another in such a way that the difference between metaphor and other non-literal tropes is illuminated.

Delineating Metaphor

The class of metaphorical expressions can be delineated by form or by function. The paradigmatic form of metaphors is the subject-predicate form *S is P*. Metaphorical expressions of this kind are the focus of, for example, John R. Searle's (1993) account of metaphor. According to Searle, a speaker utters an expression of the form *S is P* (*Juliet is the sun*) in order to convey an intended proposition (that Juliet is radiant) of the form *S is R*.

Other accounts of metaphor delineate the class of metaphorical expressions according to a specific understanding of metaphor's function to make us think of one thing as another. For example, the comparison view of metaphor is the view that metaphor functions by comparing two things (Juliet and the sun). Delineating metaphor according to this function assimilates simile to the class of metaphors.

The clear disadvantage of constructing an account of metaphor around any particular form, even paradigmatic forms, is that doing so leaves unexplained other expression forms that have a plausible claim to metaphoricity. Extended metaphors that run the length of a poem and noun-function metaphors of the form *The B of A* (*The countless gold of a merry heart*) are not easily captured by the form *S is P*. Since diverse forms can carry out the same function, functional definitions of metaphor do better at capturing non-paradigmatic forms. Functional definitions explain why we use metaphor (e.g., to compare) in a way that form alone cannot.

The Metaphorical and the Literal

Almost every metaphor theorist accepts the Deviance Thesis: metaphor is essentially nonstandard and

deviates either semantically or pragmatically from ordinary literal language. The Deviance Thesis is reflected in the persistent, but challenged (Cohen, 1976), view that metaphors are either literally false or conceptually incongruous.

Deviance and Value

Historically, the deviance of metaphor has been tied to the question of the value of metaphor. Although Cicero (*De Oratore*), Quintilian (*Institutio Oratoria*), and rhetoricians celebrated the deviance of metaphor and its embellishment of language, philosophers John Locke (*Essay concerning human understanding*) and Thomas Hobbes (*Leviathan*) condemned the use of metaphor in philosophical inquiry. If the best chance at arriving at and communicating truth is afforded only by unambiguous literal language use, metaphor's deviance from the literal is therefore suspect. Jean-Jacques Rousseau (*Essay on the origin of languages*) and Friedrich Nietzsche (*On truth and falsity in their ultra-moral sense*) attempted to undercut this criticism by arguing that all language is fundamentally metaphorical and championed metaphor's creative function. For a short history of philosophical thought about metaphor, see Johnson (1981) and Kittay (1987).

Deviance: Semantic or Pragmatic?

The cognitive value of metaphor is now generally conceded and metaphor's deviance is considered separately from its value. Contemporary accounts of metaphor characterize its deviance either as a violation of semantic rules or as a violation of pragmatic constraints. Various semantic theories can describe the semantic deviance of metaphor; for example, it can be described as a violation of selection restrictions or as a violation of standard meaning lines between possible worlds. Samuel R. Levin (1977) describes construal mechanisms for assigning interpretations to anomalous or deviant sentences, including the metaphors of ordinary language use and conversation. Metaphors are semantically deviant because they fall outside the class of sentences generated by normal operation of the rules of grammar. According to the metaphor *The stone died*, to be a stone in the metaphoric sense (to be a dunce) is to be similar in characteristics to a stone in the literal sense. The noun *stone* has semantic markers that might include (((Object) (Physical)) (Nonliving) (Mineral))), and the verb *die* has semantic markers that may include ((Process) ((Result) ((Cease to be) (Living)))). The verb has selection restrictions ((Human) or (Animal) or (Plant)), and it is these restrictions that are violated in *The stone died*. Construal rules that sanction the transfer of the feature (Human) to the semantic markers of *stone*, for example, allow us to derive the interpretation that the dunce died.

Jaakko Hintikka and Gabriel Sandu (1990) characterize the semantic deviance of metaphor according to possible world semantics. On this view, meanings are functions from possible worlds to classes of individuals. We can visualize this function by imagining that a notional meaning line connects the individuals, in their respective possible worlds, picked out by the function. Metaphoric meaning is a function that draws nonstandard or deviant meaning lines: they differ from literal meaning lines in that they rely exceptionally heavily in some specific respect on either qualitative or functional similarity. In *The stone died*, the speaker draws the meaning lines of *stone* on the basis of qualitative hardness and immovability.

Pragmatic accounts of metaphor are motivated by the observation that the very same expression (for example, *A storm is gathering*) can be interpreted literally in one context (said of a dark and windy sky) and yet be intended and interpreted metaphorically in another context (said of an anticipated faculty meeting). A full explanation of metaphor, then, must look beyond merely the expression itself to the context of the utterance. H. Paul Grice's theory of conversational implicature provides a pragmatic account of the deviance of metaphor. Conversational implicatures, including metaphor, arise when what is said violates conversational maxims. Grice (1975) says that metaphor violates the conversational maxim of saying only what is true, while Dan Sperber and Deirdre Wilson (1986) argue that it violates the principle of saying only what is relevant. In either case, it is noticing that what is said deviates from these maxims and principles that prompts the hearer to search for an interpretation of the utterance (*Juliet is the sun*) such that the speaker is contributing something true or relevant to the conversation (that Juliet chases away darkness).

Joseph Stern (2000) presents a dual challenge to the Deviance Thesis: unlike Levin, Stern argues that metaphor is not grammatically deviant and, in contrast to Gricean pragmatic accounts, that it is not necessary to first notice that what is said is deviant. Stern suggests that metaphor is a linguistic type representable in the grammar by the operator *Mthat*[Φ]. Like indexicals and demonstratives, metaphors have both a character and a content. The character, or meaning, of a metaphor is the linguistic function from context to content expressions; the content of a metaphor is the propositional content determined by the character (i.e., the interpretation of the metaphorical expression). For example, in an analysis of *Juliet is the sun*, in which the predicate is the metaphorical component, the character of the metaphor picks out the content of {*Mthat*['is the sun']}, that is, properties (nourishing, chasing away darkness, etc.) that are associated with the sun.

Stern argues that metaphor is not pragmatically deviant in the way suggested by implicature theorists, insofar as what is conveyed by a metaphor is not inferentially derived against a background of conversational maxims. Instead, the rules of grammar simultaneously make available metaphorical and literal interpretations. Stern and others (Glucksberg and Keysar, 1993; Récanati, 1995) cite evidence that the metaphorical interpretation of the sentence is processed in parallel with, and not serially to, the literal interpretation. It is not necessary to recognize first that what is said violates pragmatic constraints.

Theories of Metaphor

Conditions of Adequacy

Theories of metaphor are successful to the extent that they fulfill certain conditions of adequacy. For example, the proposal that metaphorical expressions are ambiguous fails, because it cannot explain how the meaning of the metaphorical expression depends on the literal meaning of the words used. Unlike ambiguous expressions (example, *bank*), the meaning of the words on one (literal) interpretation stay 'active' and guide the other (metaphorical) interpretation. Other desiderata include explanations of the expressive power and catachretic function of metaphor to remedy gaps in the vocabulary by using words in new ways. Accounts should make sense of the ubiquity of metaphor and explain why some metaphors fail. The more controversial features of metaphor, such as its apparent falsity and nonparaphrasability, must either be accounted for or be explained away. For further discussion of these and other conditions of adequacy, see Nogales (1999).

Aristotle

Aristotle defines metaphor as the transference of a name from genus to species, from species to genus, from species to species, or by analogy. In his influential treatment of metaphor (found in *Poetics* and in *Rhetoric*) we find the seeds of substitution, analogy, and simile theories of metaphor. Under Quintilian's substitution theory of metaphor, a new decorative name is transferred to an object in substitution for its usual plain name, though for merely rhetorical effect. Analogy theories of metaphor have been of particular interest to those interested in the predictive nature of scientific models (Hesse, 1966; Gentner, 1982). Because Aristotle thinks that the function of both simile and metaphor trades on comparison and noticing pre-existing similarities, he is also credited as an early proponent of the view that metaphor is a kind of simile. The elliptical simile theory of metaphor

specifies that metaphors such as *Time is a child at play* are ellipses of corresponding similes (*Time is like a child at play*). Lynne Tirrell (1991) argues that not all metaphors have corresponding similes, such as metaphors of the form *A is Ψ* (*The moon is envious*). Therefore, although metaphor theorists continue to be tempted to explain metaphor in terms of simile (Fogelin, 1988), elliptical simile theories of metaphor apparently cannot serve as general theories of metaphor.

Interaction Theories of Metaphor

Interaction theories of metaphor have a prominent place among contemporary semantic theories of metaphor. Introduced by I. A. Richards (1936) and Max Black (1962), this kind of theory proposes that instead of simply re-naming or comparing objects, two concepts or systems of associated commonplaces are simultaneously 'interactive.' In *my love is a rose*, the source (also called 'vehicle' or 'focus') *rose* projects an isomorphic set of connotations or believed commonplaces (such as fragility) upon the topic (also called 'tenor') *my love*. Eva Feder Kittay (1987) articulates this interaction in terms of semantic field theory. In this theory, the meaning of a term is a function of its relation (of affinity or contrast) to the other terms in its semantic or conceptual field. For example, the meaning of *midwife* is a function of its semantic field structured according to relations among the agent (midwife), her patient (the mother), and the result (the child). For Socrates's metaphor of teachers as midwives, interactive projection consists of restructuring the relations among the terms in topic field (*teacher*) analogously to those in the source field (*midwife*). Reconceiving the topic in this manner permits the special and controversial creativity of metaphor: metaphor goes beyond exploiting existing similarities to create new similarities or perspectives on the world (perspectival theory of metaphor). Metaphor makes us think of one thing as another because its function is to create a perspective from which we gain an understanding of that which is metaphorically portrayed.

Davidson and Metaphorical Meaning

Donald Davidson (1978) gives what is perhaps the most influential objection to any semantic treatment of metaphor. According to Davidson, the fundamental error of all semantic accounts of metaphor is to read the contents of the thoughts provoked by the metaphor into the content of the expression itself. Davidson denies that the concept of metaphorical meaning is required to explain how metaphor achieves its effect. Sharply distinguishing between what words

mean and what words are used to do, Davidson argues that it is the meaning of words that is supposed to explain what can be done with words (and not, for example, the effects achieved by metaphor that explain the meaning of the metaphorical expression). It is because literal meaning and truth conditions, but not metaphorical meaning, can be assigned to sentences apart from particular contexts of use that only literal meaning has genuine explanatory power. If there is no metaphorical meaning, then theories of metaphor can tell us about the effects metaphors have on us, but they do not provide a method for decoding a special content conveyed by the metaphorical expression. See Leddy (1983) and Farrell (1987) for criticisms of Davidson's view, and Crosthwaite (1985) for a defense of Davidson's account.

See also: Aristotle and Linguistics; Character versus Content; Creativity in Language; Expression Meaning versus Utterance/Speaker Meaning; Expressive Power of Language; Implicature; Irony; Maxims and Flouting; Metaphor: Psychological Aspects; Propositions; Relevance Theory; Semantics-Pragmatics Boundary; Truth Conditional Semantics and Meaning.

Bibliography

- Beardsley M (1962). 'The metaphorical twist.' *Philosophy and Phenomenological Research* 22, 293–307.
- Berg J (1999). 'Referential attribution.' *Philosophical Studies* 96, 73–86.
- Black M (1962). *Models and metaphor*. Ithaca: Cornell University Press.
- Cohen T (1976). 'Notes on metaphor.' *Journal of Aesthetics and Art Criticism* 34, 249–259.
- Cooper D E (1986). *Metaphor*. Oxford: Blackwell.
- Crosthwaite J (1985). 'The meaning of metaphors.' *Australasian Journal of Philosophy* 63, 320–335.
- Davidson D (1978). 'What metaphors mean.' In Sacks S (ed.) *On metaphor*. Chicago: University of Chicago Press.
- Farrell F B (1987). 'Metaphor and Davidsonian theories of meaning.' *Canadian Journal of Philosophy* 17, 625–642.
- Fogelin R (1988). *Figuratively speaking*. New Haven: Yale University Press.
- Gentner D (1982). 'Are scientific analogies metaphors?' In Miall D S (ed.) *Metaphor: problems and perspectives*. Atlantic Highlands: Humanities Press Inc. 106–132.
- Glucksberg S & Keysar B (1993). 'How metaphors work.' In Ortony A (ed.) *Metaphor and thought*, 2nd edn. New York: Cambridge University Press. 401–424.
- Goodman N (1968). *Languages of art*. Indianapolis: Bobbs-Merrill.
- Grice H P (1975). 'Logic and conversation.' In Cole P & Morgan J L (eds.) *Syntax and semantics 3: speech acts*. New York: Academic Press. 41–58.
- Hesse M B (1966). *Models and analogies in science*. Notre Dame: University of Notre Dame Press.

- Hintikka J (ed.) (1994). *Aspects of metaphor*. Dordrecht: Kluwer Academic Publishers.
- Hintikka J & Sandu G (1990). 'Metaphor and the varieties of lexical meaning.' *Dialectica* 44, 55–78.
- Johnson M (ed.) (1981). *Philosophical perspectives on metaphor*. Minneapolis: University of Minnesota Press.
- Kittay E F (1987). *Metaphor: its cognitive force and linguistic structure*. New York: Oxford University Press.
- Lakoff G & Johnson M (1980). *Metaphors we live by*. Chicago: The University of Chicago Press.
- Leddy T (1983). 'Davidson's rejection of metaphorical meaning.' *Philosophy and Rhetoric* 16, 63–78.
- Levin S R (1977). *The semantics of metaphor*. Baltimore: The Johns Hopkins University Press.
- Mac Cormac E R (1985). *A cognitive theory of metaphor*. Cambridge: MIT Press.
- Moran R (1997). 'Metaphor.' In Hale B & Wright C (eds.) *A companion to the philosophy of language*. Oxford: Blackwell Publishers. 248–268.
- Nogales P D (1999). *Metaphorically speaking*. Stanford: CSLI Publications.
- Radman Z (ed.) (1995). *From a metaphorical point of view: a multidisciplinary approach to the cognitive content of metaphor*. Berlin: de Gruyter.
- Récanati F (1995). 'The alleged priority of literal interpretation.' *Cognitive Science* 19, 207–232.
- Richards I A (1936). *The philosophy of rhetoric*. Oxford: Oxford University Press.
- Searle J R (1993). 'Metaphor.' In Ortony A (ed.) *Metaphor and thought*, 2nd edn. New York: Cambridge University Press. 83–111.
- Sperber D & Wilson D (1986). *Relevance: communication and cognition*. Cambridge: Harvard University Press.
- Stern J (2000). *Metaphor in context*. Cambridge: MIT Press.
- Tirrell L (1991). 'Reductive and nonreductive simile theories of metaphor.' *The Journal of Philosophy* 88, 337–358.

Metaphor: Psychological Aspects

R Gibbs, University of California, Santa Cruz, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

The psychological study of metaphor has had a major impact on the interdisciplinary understanding of language and thought. Thirty years ago, the topic of metaphor was mostly seen as peripheral to the major focus of research in both linguistics and psychology, because metaphor was primarily viewed as a poetic device that is not representative of how people ordinarily speak or think. But in conjunction with the emergence of cognitive linguistics in the 1970s and 1980s, psychological research has demonstrated that metaphor is ubiquitous in discourse, can often be easily understood and produced in appropriate social and linguistic contexts, and perhaps most importantly, is both a type of language use and a fundamental scheme of thought. This entry describes the empirical evidence relevant for, and the theories building on, these claims.

The Ubiquity of Metaphor in Language

Metaphor has traditionally been viewed as a distortion of both thought and language, because it involves the transfer of a name to some object to which that name does not properly belong. Speakers and writers presumably use metaphor as an ornamental feature for poetic and rhetorical purposes (e.g., to say what is difficult to state literally, to express meaning in a vivid manner), rather than to impart fundamental

concepts. In each case of metaphorical language, a person aims to present some underlying analogy or similarity in the form of a condensed or elliptical simile. Thus, a metaphor of the 'A is B' form indirectly implies the speaker's intended literal meaning "A is like B in certain respects." For instance, the metaphor 'The car beetles along the road' describes the movement of the car as being like the movement of a beetle. Under this traditional view, metaphor should be infrequent in language, especially in scientific discourse, and people should have more cognitive difficulty when uttering and understanding metaphors than they do when using the equivalent literal speech.

Psychological research has shown, however, that metaphor is a major part of both spoken and written language. Various studies have attempted to quantify the frequency of metaphor use in a variety of contexts. One detailed study examined the use of metaphor in transcripts of psychotherapeutic interviews, in various essays, and in the 1960 Kennedy-Nixon presidential debates and found that people used 1.80 novel and 4.08 frozen metaphors per minute of discourse (Pollio *et al.*, 1977). If one assumes that people engage in conversation for as little as 2 hours per day, a person would utter 4.7 million novel and 21.4 million frozen metaphors over a 60-year life span! A different analysis of the metaphors produced in television debates and news commentary programs showed that speakers use one unique metaphor for every 25 words (Graesser *et al.*, 1989). These, admittedly crude, analyses clearly demonstrate that metaphor is not the special privilege of a few gifted

speakers, but is ubiquitous throughout both written and spoken discourse.

However, a closer look at everyday language suggests that these empirical attempts to 'count' instances of metaphor vastly underestimate the pervasiveness of metaphor in people's ordinary speech. Typical frequency counts of metaphor do not include analysis of conventional speech that is motivated by metaphoric modes of thought. Consider the following mundane expressions that people often use in talking about verbal arguments (Lakoff and Johnson, 1980).

Your claims are indefensible.
I've never won an argument with him.
I demolished his argument.
He attacked every weak point in my argument.
His criticisms were right on target.
He shot down all of my arguments.

At first glance, none of these expressions appear to be very metaphoric, at least in the same way that an utterances such as 'The sun is the eye of heaven' might be. Yet, a closer look reveals the systematic metaphoric structuring whereby people think of arguments in terms of wars. We can actually win or lose arguments. We see the person we are arguing with as an opponent. We attack his positions, and we defend our own. We plan and use strategies. We might find certain positions undefensible, requiring us to take new lines of attack. Each of these things do not simply reflect the way we talk about arguments: we actually argue as if we were in a war. Our understanding of argument as war is active and widespread, but this concept is so deeply entrenched in our ordinary conceptual system that we tend to miss its metaphorical character.

Cognitive linguistic research has suggested that there are perhaps hundreds of conceptual metaphors, such as ARGUMENTS ARE WARS, that structure our everyday experience, and that they are found in a wide variety of conceptual domains (Gibbs and Steen, 1999; Kövecses, 2002; Lakoff and Johnson, 1999). Linguistic analyses do not typically quantify the number of verbal metaphors, and the conceptual metaphors underlying them, that may be present in any one sample of speech or text. But one psychological study of the narratives women produced when describing their experiences recovering from cancer showed that conventional metaphoric language was employed more than 6 times per minute, and that only 22 conceptual metaphors underlay the vastly different metaphoric expressions these women produced, especially in their talk of emotion (Gibbs and Franks, 2002). Conceptual metaphors seem to be ubiquitous in the ways people talk of their experiences.

One question that has generated a great deal of debate within psychology is whether these instances of conventional metaphor necessarily reflect anything about the metaphorical nature of many abstract concepts.

Metaphor Understanding: The Standard View

The traditional belief that metaphor is deviant suggests that metaphors should be more difficult to interpret than literal speech. The most famous proposal along this line comes from H. Paul Grice's theory of conversational implicature (Grice, 1989) (also see Grice, H. Paul (1913–1988); *see* **Implicature**). Grice argued that the inferences needed to understand nonliteral meaning are derived from certain general principles or maxims of conversation that participants in talk-exchange are mutually expected to observe (Grice, 1989) (*see* **Maxims and Flouting**). Among these are expectations that speakers are to be informative, truthful, relevant, and clear in what they say. When an utterance appears to violate any of these maxims, as in the case of metaphor, listeners are expected to derive an appropriate 'conversational implicature' about what the speaker intended to communicate in context, given the assumption that he or she is trying to be cooperative (*see* **Cooperative Principle**).

Grice (1989) more specifically suggested what has become known as the 'standard pragmatic model' for understanding indirect and nonliteral meanings, including metaphor. In this view, understanding metaphor is accomplished in a series of steps: (1) analyze the literal meaning of an entire expression, (2) compare this literal meaning to the context, (3) if the literal meaning is appropriate, then stop, otherwise (4) derive an alternative meaning that makes the speaker's/writer's utterance sensible in the context, given the cooperative principle. This rational account suggests, then, that metaphors are understood as conversational implicatures and should take additional time to comprehend over that needed to interpret literal speech that is appropriate to the context.

Psychological Tests of the Standard View

How accurate is the standard view as a psychological theory of metaphor understanding? First, the results of many reading-time experiments in psycholinguistics show that people do not always require additional mental effort to comprehend many kinds of figurative utterances, as compared with so-called literal speech (Gibbs, 1994, 2002). Listeners/readers often take no longer to understand the figurative

interpretations of metaphor (e.g., 'billboards are warts on the landscape'), metonymy (e.g., 'The ham sandwich left without paying'), sarcasm (e.g., 'You are a fine friend'), idioms (e.g., 'John popped the question to Mary'), proverbs (e.g., 'The early bird catches the worm'), and indirect speech acts (e.g., 'Would you mind lending me five dollars?') than to understand equivalent literal expressions, particularly if these are seen in realistic linguistic and social contexts. Appropriate contextual information provides a pragmatic framework for people to understand metaphoric utterances without any recognition that these utterances violate conversational norms. In fact, psychological studies have specifically shown that people do not need to find a defective literal meaning before searching for a nonliteral meaning. For example, people apprehend the metaphoric meanings of simple comparison statements (e.g., 'surgeons are butchers') even when the literal meanings of these statements fit perfectly with the context (Glucksberg *et al.*, 1982). Even without a defective literal meaning to trigger a search for an alternative meaning, metaphor can be automatically interpreted.

These experimental findings from psycholinguistics are damaging to the general assumption that people understand metaphor as violations of conversational maxims. Similar psychological mechanisms appear to drive the understanding of both literal and metaphoric speech, at least insofar as early cognitive processes are concerned. Everyone agrees that people may sometimes take a good deal of time to process novel poetic metaphors, for example. Studies have shown, in fact, that conventional, or familiar, metaphors can be understood more quickly than novel expressions (Katz and Ferretti, 2001). Yet the additional time needed to understand novel metaphors is not necessarily due to a preliminary stage during which the literal meaning for an entire utterance is first analyzed and then rejected. Listeners may take longer to understand a novel expression, such as 'The night sky was filled with molten silver,' because of the difficulty in integrating the figurative meaning with the context, and not because listeners are first analyzing and then rejecting the expression's literal meaning (Schraw, 1995).

Many psychologists have gone on to argue that even if metaphor does not necessarily demand extra cognitive effort to understand, people may still analyze literal, conventional, or salient, aspects of word meaning during immediate metaphor comprehension (Blasko and Connine, 1993; Giora, 2001). Some studies, which measure the meanings activated during each part of the moment-by-moment process of linguistic understanding, suggest that comprehending familiar and novel metaphors engages different

linguistic processes. Analysis of literal word meaning still precedes metaphorical meaning during novel metaphor understanding, with both types of meaning arising in parallel during familiar metaphor processing. Other studies that assessed people's speeded judgments about the sensibility of different word strings at different moments find no difference in the comprehension speed for literal and figurative strings (McElree and Nordlie, 1999). This lack of time-course differences is inconsistent with the claim that metaphoric interpretations are computed after a literal meaning has been analyzed, and suggest that literal and metaphoric interpretations are computed in parallel.

Although these research findings imply that metaphor processing is not secondary to literal understanding, psycholinguists are, again, careful to note that people may be biased toward initially interpreting the literal, or salient, meanings of metaphoric statements in cases of novel metaphor (Giora, 2001). Yet others argue that even if some linguistic meanings (e.g., literal or metaphoric) are created sooner during metaphor processing, these findings do not imply that entirely different mental processes operate to produce these different meanings (Gibbs, 2002). Different kinds of meaning may arise from a single linguistic process. The fact that scholars label one kind of meaning 'literal' and another 'metaphoric' doesn't necessarily indicate that different processes operate (such as a literal processing mode and a metaphoric processing mode) as people access these meanings (either in a serial or parallel manner).

More recent theories of figurative language understanding, which are more general than metaphor theories *per se*, suggest that people may initially access a word's interpretation that can be compatible with both its literal and metaphoric meanings (Frisson and Pickering, 2001). Over time, however, people use context to home in on the word's appropriate metaphoric meaning, where the homing-in process is faster when the preceding context is strong, and slower when the preceding context is neutral. In this way, context does not operate to distinguish between different literal and metaphoric meanings, as assumed by most theories (such as in the standard model), but functions to change an underspecified, or highly general meaning, into a contextually appropriate, specific interpretation which may be metaphorical.

A different theory embraces the notion of 'constraint satisfaction' to provide a comprehensive model of the different sources of information that constrain metaphor understanding (Katz and Ferretti, 2001). Under this view, understanding a metaphoric utterance requires people to consider different linguistic (e.g., people's familiarity with words and phrases) and nonlinguistic (e.g., related to specific

context) information that best fits together to make sense of what a speaker or writer is saying. These different sources of information are probabilistically evaluated, and combined to offer a most likely 'winning' meaning for a metaphor. A constraint satisfaction model may have the flexibility to account for a wide variety of metaphor processing data that seems to differ depending on the familiarity or conventionality of the expression, the context in which it is encountered, and the speaker's/writer's likely intentions in using metaphorical language.

In summary, there has been a great deal of psychological research devoted to the general question of whether metaphorical language requires additional cognitive effort to understand, compared to non-metaphorical speech. The findings of these widely varying studies strongly imply that metaphors are not deviant and do not necessarily take more time to understand, but that more subtle factors, such as the familiarity of the expression and the context in which it is used, can shape the time-course of metaphor understanding. Many studies now situate metaphor understanding within a more comprehensive view of linguistic processing that does not posit specialized mechanisms for interpreting metaphors, even if these expressions often convey distinctive kinds of meanings (Kintsch and Bowles, 2002), and which specifically relies on cognitive mechanisms, such as suppression, that are employed widely in all aspects of language processing (Gernsbacher and Robertson, 1999).

Psychological Models of Metaphor Understanding

A great deal of research has been devoted to the specific processes involved in understanding metaphorical meaning, beyond the general question of whether metaphors are more difficult to comprehend than literal speech. These studies have explicitly examined the ways that the A, or target, and B, or vehicle, terms interact to produce metaphorical meaning. A long-standing assumption in many academic fields is that we understand metaphors by recognizing the ways that topic and vehicle terms are similar. Thus, in understanding the metaphor 'Juliet is the sun,' listeners are presumed to figure out the properties of both Juliet and the sun that are similar.

But psychological studies indicate that metaphor comprehension does not demand that the topic and vehicle terms share properties or associations (Camac and Glucksberg, 1984). This finding is supported by many studies showing that metaphors have directional meaning. If metaphorical meaning arises from the overlap of the semantic features of topic and vehicle, expressions such as 'The surgeon is a butcher' and

'The butcher is a surgeon' should have similar metaphorical meanings. But this is clearly not the case. The similarity that arises from the comparison of a topic and vehicle does not produce metaphorical meaning. Instead, similarity is created as an emergent property of understanding metaphor. Thus, many psychological studies have demonstrated that novel features emerge from metaphor comprehension that are not salient in one's separate understanding of the topic or vehicle (Gineste *et al.*, 2000). This idea is consistent with the popular, but somewhat vague, interactionist theory of metaphor (Black, 1979), which argues that the presence of the topic stimulates a listener to select one of the vehicle's properties so as to construct a 'parallel implication complex' that may induce changes in one's understanding of both the topic and vehicle. In general, psychological studies provide strong evidence supporting the idea that metaphor cannot be reduced to rule-governed extensions or variations of the topic's and vehicle's literal meanings.

Psychologists disagree, however, about the cognitive mechanisms involved in feature emergence during metaphor understanding. The two main proposals state that metaphorical mappings between concepts from dissimilar domains can be accomplished by either comparison or categorization processes. Traditional comparison theories posit that metaphor understanding demands a mapping of low-salient features from the source domain with high-salient features of the target domain (Miller, 1979). But understanding many metaphors, such as 'Men are wolves,' seems to involve the activation of semantic features that are not typically associated with either the source or target domain until after the metaphor has been understood (Ortony, 1979). Gentner's 'structure-mapping' theory of analogy and metaphor avoids this problem by suggesting that people begin processing a metaphor by first aligning the representations of the source and target domain concepts (see Gentner *et al.*, 2001). Once these two domains are aligned, further inferences are directionally projected from the source to the target domain. Finally, new inferences arise within the target domain, reflecting relational, and not just feature-specific, aspects of the metaphor comprehension processes. Experimental evidence in support of this comparison view shows, for instance, that people infer relational, but not feature-specific, meanings when interpreting metaphors (Gentner *et al.*, 2001). For instance, when people read 'Plant stems are drinking straws,' they infer that both plants and straws convey liquid to nourish living things, and not just that both plants and straws are long and thin (i.e., object commonalities). Other research indicated that metaphors that express relational information (e.g., 'Plant stems are drinking straws')

are viewed as being far more apt than those that only map object features ('Her arms were like twin swans').

An alternative view claims that metaphors are better understood via categorization processes, as class-inclusion, rather than comparison, statements (Glucksberg, 2001). For example, the statement 'Yeltsin was a walking time bomb' asserts that the former Russian President is a member of a category that is best exemplified by time bombs. Of course, time bombs can belong to several other categories, such as the weapons used by terrorists. But in the context of talking about people, time bombs best exemplify the abstract category of 'things that explode at some unpredictable time in the future and cause a lot of damage.' In this way, metaphors reflect *ad hoc* categories and refer at two levels: the concrete level (i.e., an explosive device) and a superordinate level (i.e., the properties of time bombs).

One implication of the class-inclusion model is that it suggests that the topics and vehicles, or target and source domains, in metaphors play different but interactive roles in metaphor comprehension. For example, the word 'snake' evokes different meanings in the phrases 'my lawyer is a snake' and 'the road was a snake.' In this way, metaphor topics provide dimensions for attribution, while vehicles provide properties to be attributed to the topic. Psychological evidence supporting this position showed that in a reading-time study, presenting people first with a topic term that is highly constrained reduces the time needed for the subsequent processing of a metaphorical statement, in contrast to when people are first presented with a less-constrained topic (Glucksberg, 2001). Furthermore, presenting people with an unambiguous vehicle primes subsequent metaphor comprehension, in contrast to what happens when they are presented with an ambiguous vehicle term. This pattern of data illustrates how the level of constraint is an important feature of metaphor topics, while the degree of ambiguity is an important characteristic of metaphor vehicles. Comparison models of metaphor understanding are unable to explain the importance of constraint and ambiguity, because they assume that metaphor comprehension always begins with an exhaustive extraction of the properties associated with both topics and vehicles. Having advance knowledge about either the topic or vehicle should presumably, then, prime metaphor processing. However, the categorization view correctly predicts that only advanced knowledge about highly constrained topics and unambiguous vehicles facilitates metaphor comprehension, a finding that is most consistent with the claim that metaphor understanding involves creating a new, *ad hoc* category and not merely comparing one's knowledge about topic and vehicle domains.

A proposal titled the 'career of metaphor' theory combines aspects of both the comparison and categorization views (Gentner and Bowdle, 2001). This theory claims that there is a shift in the mode of mappings from comparison to categorization processes as metaphors become conventionalized. For instance, novel metaphors such as 'Science is a glacier' involve base terms, such as 'glacier,' with a literal source (i.e., 'a large body of ice spreading outward over a land surface'), but no relational metaphoric sense (i.e., 'anything that progresses slowly but steadily'). People comprehend novel metaphors as comparisons in which the target concept (e.g., 'science') must be structurally aligned with the literal base concept (e.g., 'glacier'). In some instances, the comparison process may lead to the induction of a novel metaphor category. On the other hand, conventional metaphors can be understood either by comparison or categorization processes. For example, the metaphor 'A gene is a blueprint' has two closely related senses (e.g., 'a blue and white photographic print detailing an architect's plans' and 'anything that provides a plan'). The relations between these two senses make the conventional base term polysemous (i.e., semantically related literal and metaphoric meanings). As such, conventional metaphors may be understood by matching the target concept with the literal base concept (a comparison process) or by viewing the target concept as a member of the superordinate metaphoric category named by the base term (a categorization process).

Metaphor in Thought

Most of the psychological research on metaphor has focused on how it is used and understood within language, and has assumed that metaphorical meaning is created *de novo*, and does not reflect preexisting aspects of how people ordinarily conceptualize ideas and events in terms of pervasive metaphorical schemes. But in the past 20 years, various linguists, philosophers, and psychologists have embraced the alternative possibility that metaphor is fundamental to language, thought, and experience. Cognitive linguists, for instance, claim that metaphor is not merely a figure of speech, but is a specific mental and neural mapping that influences a good deal of how people think, reason, and imagine in everyday life (Lakoff and Johnson, 1999). Evidence supporting this claim comes from linguistic research on the historical evolution of what words and expressions mean, the systematicity of conventional expressions within and across languages, novel extensions of conventional metaphors, studies on polysemous word meaning, and nonverbal behaviors such as gesture (Gibbs, 1994; Lakoff and Johnson, 1980, 1999).

However, psychologists have been critical of much of this work and its possible implications for theories about conceptual structure and metaphor understanding. First, most of the evidence for metaphorical thought, or conceptual metaphor, comes from purely linguistic analyses, and psychologists have expressed deep skepticism about these claims on both methodological and theoretical grounds, especially with regard to linguists' heavy reliance on their own linguistic intuitions (Murphy, 1996). Second, some psychologists argue that conceptual metaphor theory is unfalsifiable if the only data in its favor is the systematic grouping of metaphors linked by a common theme (Vervaeke and Kennedy, 1996). Consider again the conceptual metaphor ARGUMENT IS WAR (Lakoff and Johnson, 1980), which presumably motivates conventional expressions such as 'He attacked my argument' and 'He defended his position.' Cognitive linguistic research suggests that any expression about argument that does not fit the WAR theme is usually seen as evidence for another theme, such as WEIGHING, TESTING, or COMPARING. This implies that no linguistic statement can be brought forward as evidence against the ARGUMENT IS WAR metaphor, which makes the basic tenet of conceptual metaphor theory impossible to falsify. Finally, some psychologists argue that many conventional expressions viewed as metaphorical by cognitive linguists are really not metaphorical at all, but are treated by ordinary speakers/listeners as literal speech (Glucksberg, 2001). Simple expressions like 'He was depressed' are entirely literal, and may not be motivated by a conceptual metaphor such as SAD IS DOWN, because they only reflect something about the polysemous nature of meaning (e.g., 'depression' can be used to talk about either physical depression or emotional depression).

Psychological Studies on Conceptual Metaphor

Despite the skeptical reaction of some psychologists to the idea of metaphorical thought, or conceptual metaphor, there is a great deal of psychological evidence supporting the claim that many aspects of people's abstract concepts and reasoning processes are shaped by enduring conceptual metaphor. Studies show, for instance, that conceptual metaphors influence the ways people conceive of various abstract domains, such as emotions, minds, politics, advertising, scientific theories, the self, morality, learning, and problem-solving (Gibbs, 1994; see Steen and Gibbs, forthcoming, for reviews). Most of these studies demonstrate that providing people with a particular metaphorical construal of some domain (e.g., that EMOTIONS ARE

CONTAINERS) can facilitate the way they learn new information, solve problems, and make decisions, if the newly encountered material has a similar metaphorical structure. At the same time, whereas switching from one conceptual metaphor to another may require more cognitive effort in some situations (Langston, 2002), people typically have multiple metaphorical ways of conceiving of most abstract ideas (e.g., THEORIES ARE BUILDINGS, THEORIES ARE FABRIC (Gibbs, 1994). This multiplicity of metaphorical schemes provides another source of evidence for the idea that a good deal of ordinary thought is shaped by metaphor.

Even if people seem able to think metaphorically about various domains, many psychologists and especially many psycholinguists are skeptical about whether conceptual metaphors are normally recruited during people's ordinary comprehension of language (Glucksberg, 2001). These critics find it difficult to believe that conceptual metaphors play much of a role in how people interpret verbal metaphors such as 'Surgeons are butchers' or 'Lawyers are snakes.' To a large extent, the debate over conceptual metaphor settles into two camps: those scholars studying novel metaphors and those studying conventional language that may reflect different conceptual metaphors (e.g., 'He attacked my argument' for ARGUMENTS ARE WARS, 'Our relationship hit a dead end street' for LIFE IS A JOURNEY, and so on). Thus, different approaches to the psychology of metaphor understanding are oriented toward different types of metaphorical language. A likely possibility is that conceptual metaphor may have a strong influence on some aspects of verbal metaphor use, but not on others.

In fact, there is a large body of evidence from psychological studies, employing different methods, that clearly demonstrates that (a) people conceptualize certain topics via metaphor, (b) conceptual metaphors assist people in tacitly understanding why metaphorical words and expressions mean what they do, and (c) people access conceptual metaphors during their immediate, online production and comprehension of conventional and novel metaphors. This work includes studies investigating people's mental imagery for conventional metaphors, as in idioms and proverbs (Gibbs and O'Brien, 1990), people's context-sensitive judgments about the figurative meanings of idioms in context (Nayak and Gibbs, 1990), people's immediate processing of idioms (Gibbs *et al.*, 1997), people's responses to questions about metaphorical expressions about time (Boroditsky and Ramscar, 2002; Gentner *et al.*, 2002), readers' understanding of metaphorical time expressions (McGlone and Harding, 1998), and studies looking at the embodied foundation for conventional metaphoric language (Gibbs *et al.*, 2004).

To briefly give a few examples from these psycholinguistic experiments, studies show that people have a complex metaphorical understanding of many abstract domains, which partially motivates everyday reasoning and language use. For instance, people conceive of the domain of emotions metaphorically, based partly on their embodied experiences of emotions, such that they tacitly know that phrases like 'blow your stack' and 'flip your lid' are motivated by the conceptual metaphor of ANGER IS HEATED FLUID IN A CONTAINER. This metaphorical understanding of anger influences people's judgments about the degree to which someone experiences anger and about the best use of different metaphorical phrases in context (Nayak and Gibbs, 1990). At the same time, people's tacit knowledge of conceptual metaphors constrains the specific mental images they can form for verbal metaphors, and the specific meanings they believe these metaphors express (e.g., that 'blow your stack' means to get very angry while the person is feeling internal pressure, and the expression of the anger is unintentional and forceful) (Gibbs and O'Brien, 1990). In fact, many conventional phrases and idioms, long thought to be dead metaphors, retain much of their metaphorical meaning precisely because they continue to be linked to enduring conceptual metaphors. Finally, priming studies suggest that reading a conventional metaphorical phrase, such as 'John blew his stack,' quickly accesses the conceptual metaphor (ANGER IS HEATED FLUID IN A CONTAINER) that partly motivates why this expression has the particular metaphorical meaning it conveys (Gibbs *et al.*, 1997). Reading another expression with roughly similar metaphoric meaning, such as 'John bit her head off,' activates a different conceptual metaphor (ANGER IS ANIMAL BEHAVIOR), giving rise to the creation of these metaphorical expressions.

The debate over the role that metaphorical thought may play in a psychological theory of verbal metaphor use will likely continue. Once more, it seems inevitable that several of the different approaches to metaphor within linguistics and psychology will become part of a more comprehensive theory of metaphor. Yet it is already evident that the traditional views of metaphor as deviant, ornamental aspects of language and thought no longer are tenable and that psychological studies have provided excellent reasons to believe that metaphor is a fundamental part of the ways people speak and think.

See also: Aristotle and Linguistics; Cooperative Principle; Implicature; Maxims and Flouting; Metaphor: Philosophical Theories.

Bibliography

- Black M (1979). 'More on metaphor.' In Ortony A (ed.). 1–18.
- Blasko D & Connine C (1993). 'Effects of familiarity and aptness on metaphor processing.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19, 295–308.
- Boroditsky L & Ramscar M (2002). 'The roles of body and mind in abstract thought.' *Psychological Science* 13, 185–189.
- Camac M & Glucksberg S (1984). 'Metaphors do not use associations between concepts, they create them.' *Journal of Psycholinguistic Research* 13, 443–445.
- Frisson S & Pickering M (2001). 'Obtaining a figurative interpretation of a word: Support for underspecification.' *Metaphor and Symbol* 16, 149–172.
- Gentner D & Bowdle B (2001). 'Convention, form, and figurative language processing.' *Metaphor and Symbol* 16, 223–248.
- Gentner D, Bowdle B, Wolff P & Boronat C (2001). 'Metaphor is like analogy.' In Gentner D, Holyoke K & Kokinov B (eds.) *The analogical mind: perspectives from cognitive science*. Cambridge, MA: MIT Press. 199–253.
- Gentner D, Imai M & Boroditsky L (2002). 'As time goes by: understanding time as spatial metaphor.' *Language and Cognitive Processes* 17, 537–565.
- Gernsbacher M & Robertson R (1999). 'The role of suppression in figurative language comprehension.' *Journal of Pragmatics* 31, 1619–1630.
- Gibbs R (1994). *The poetics of mind: figurative thought, language, and understanding*. New York: Cambridge University Press.
- Gibbs R (2002). 'A new look at literal meaning in understanding what speakers say and implicate.' *Journal of Pragmatics* 34, 457–486.
- Gibbs R & Franks H (2002). 'Embodied metaphor in women's narratives about their experiences with cancer.' *Health Communication* 14, 139–166.
- Gibbs R & O'Brien J (1990). 'Idioms and mental imagery: the metaphorical motivation for idiomatic meaning.' *Cognition* 36, 35–64.
- Gibbs R & Steen G (eds.) (1999). *Metaphor in cognitive linguistics*. Amsterdam: Benjamins.
- Gibbs R, Cosia Lima P & Françoze E (2004). 'Metaphor is grounded in embodied experience.' *Journal of Pragmatics* 36, 1189–1210.
- Gibbs R, Bogdonovich J, Sykes J & Barr D (1997). 'Metaphor in idiom comprehension.' *Journal of Memory and Language* 37, 141–154.
- Gineste M-D, Indurkha B & Scart V (2000). 'Emergence of features in metaphor comprehension.' *Metaphor and Symbol* 15, 117–136.
- Giora R (2001). *On our mind: salience and context in figurative language understanding*. New York: Oxford University Press.
- Glucksberg S (2001). *Understanding figurative language*. New York: Oxford University Press.
- Glucksberg S, Gildea P & Bookin H (1982). 'On understanding nonliteral speech: can people ignore

- metaphors?' *Journal of Verbal Learning and Verbal Behavior* 21, 85–98.
- Graesser A, Mio J & Mills K (1989). 'Metaphors in persuasive communication.' In Meutsch D & Viehoff R (eds.) *Comprehension and literary discourse*. Berlin: De Gruyter. 131–154.
- Grice H P (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.
- Katz A & Ferretti T (2001). 'Moment-by-moment reading of proverbs in literal and nonliteral contexts.' *Metaphor and Symbol* 16, 193–222.
- Kintsch W & Bowles A (2002). 'Metaphor comprehension: what makes a metaphor difficult to understand?' *Metaphor and Symbol* 17, 249–262.
- Kövecses Z (2002). *Metaphor: a students' introduction*. New York: Oxford University Press.
- Lakoff G & Johnson M (1980). *Metaphors we live by*. Chicago: Chicago University Press.
- Lakoff G & Johnson M (1999). *Philosophy in the flesh*. New York: Basic Books.
- Langston W (2002). 'Violating orientational metaphors slows reading.' *Discourse Processes* 34, 281–310.
- McElree B & Nordlie J (1999). 'Literal and figurative interpretations are computed in equal time.' *Psychonomic Bulletin and Review* 6, 486–494.
- McGlone M & Harding J (1998). 'Back (or forward) to the future: The role of perspective in temporal language comprehension.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, 1211–1223.
- Miller G (1979). 'Images and models, similes and metaphors.' In Ortony A (ed.). 203–253.
- Murphy G (1996). 'On metaphoric representation.' *Cognition* 60, 173–186.
- Nayak N & Gibbs R (1990). 'Conceptual knowledge in the interpretation of idioms.' *Journal of Experimental Psychology: General* 119, 315–330.
- Ortony A (ed.) (1979a). *Metaphor and thought*. New York: Cambridge University Press.
- Ortony A (1979b). 'Beyond literal similarity.' *Psychological Review* 86, 161–180.
- Pollio H, Barlow J, Fine H & Pollio M (1977). *Psychology and the poetics of growth: figurative language in psychology, psychotherapy, and education*. Hillsdale, NJ: Erlbaum.
- Schraw G (1995). 'Components of metaphoric processes.' *Journal of Psycholinguistic Research* 24, 23–38.
- Steen G & Gibbs R (forthcoming). *Finding metaphor in language and thought*. Amsterdam: Benjamins.
- Vervaeke J & Kennedy J (1996). 'Metaphor in language and thought: falsification and multiple meanings.' *Metaphor and Symbolic Activity* 11, 273–284.

Metaphysics, Substitution *Salva Veritate* and the Slingshot Argument

R J Stainton, University of Western Ontario,
London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Metaphysics and Language: Facts, Propositions and 'MCT Operators'

There are lots of examples of metaphysical conclusions being drawn on the basis of linguistic phenomena. Philosophers have tried to prove the metaphysical conclusion that a benevolent all-powerful God exists on the basis of claims about what 'God' means. Others have looked at tense in language and drawn conclusions about the metaphysics of time. This article presents a more technical kind of linguistically based argument, about the metaphysics of facts, propositions and modal/causal/temporal properties – an argument whose claims are especially important in the present context because of their implications for linguistic semantics.

The article is structured as follows. In this section, I first introduce two key metaphysical views: about facts and about factlike propositions. I then introduce, also in this section, some seemingly plausible hypotheses about modality, causation, and temporal order. These are the

metaphysical views which will be attacked on linguistic grounds. In the next section I explain what substitutivity principles are, and survey three different kinds of substitutivity and nonsubstitutivity. Such principles clarify the linguistic phenomena that will be used to attack the metaphysical views. The first two sections essentially introduce crucial background material for the argument. In the third section, I then describe the modest means deployed in attacking the metaphysical targets: the slingshot argument. I end by noting some standard replies to this language-based argument.

I begin with facts. It seems natural enough to think of the world as not being exhausted by the objects in it. That is, to list all the objects in the world is not to say all there is to say about it: one also wants to know what properties the objects have, and what relations they stand in, etc. And to say all that, is seemingly to describe facts. It's also tempting to think that facts consist of complexes of ordinary, everyday objects, properties, relations, etc.; to invent a label, it is tempting to think that facts are 'world bound.' The first view that the slingshot calls into question, however, is precisely that there are such world-bound facts, in the plural. Instead, the argument seeks to show that, if facts are

made up of commonplace worldly things, there can be at most one ‘giant’ one. To be clear about the metaphysical target, it is not the very idea of facts that would have to go, for there are alternative accounts of facts which, at first glance anyway, are not threatened by the slingshot. But those accounts of facts face difficulties of their own – difficulties which world-bound facts, facts as complexes of ordinary things, are not subject to. First, as will emerge below, a rose by any other name would smell as sweet. This suggests that the fact that the rose smells sweet has the rose itself in it, not some way of linguistically describing the rose. Second, part of the point of introducing facts and the like is to see how something distinct from a sentence can serve as the sentence’s meaning, hence, we can’t have every linguistic change in the sentence giving rise to a different fact (or factlike thing). (Strawson, 1950 made this point early on.) Third, we want the same fact (or factlike thing) to give the meaning of sentences in different languages, which again means that a mere change in the word used ought not change the proposition denoted – otherwise ‘It’s raining’ and ‘Está lloviendo’ could end up denoting different things. (I’ll return to this at the end of the article.) All of these points suggest that facts are world bound. But, as we shall see, reflection upon language – and especially upon substitutivity principles – makes it hard to see how they could be.

This takes us to a related point. Positing facts often goes along with the idea that factlike things can serve as the denotations of sentences. For instance, ‘Ottawa is the capital of Canada’ might be taken to stand for something like the fact that Ottawa is the capital of Canada. And ‘Abe Lincoln was assassinated’ might be taken to denote the fact that Abe Lincoln was assassinated. Of course, it can’t really be facts *per se* that are sentence denotations, because false sentences are meaningful, and they cannot stand for facts. (For example, what fact could ‘Toronto is the capital of Canada’ stand for, given that Toronto is not the capital? There simply is no such fact.) Still, something factlike might do the trick: sentence denotations could be taken to be propositions. But the slingshot argument, in attacking facts in the plural, equally calls into question the idea that sentences stand for factlike things: if the argument succeeds, there are no such things, facts or propositions, in the plural, that sentences could denote. Again, there can be at most one thing denoted by all true sentences. And, patently, no one thing may serve as the meaning of such diverse sentences as ‘Abe Lincoln was assassinated,’ ‘Ottawa is the capital of Canada,’ ‘Russia is bigger than Uruguay,’ etc.

So, one metaphysical idea is that facts are complexes built from commonplace objects, properties and relations. Another is that propositions exist

independently of language, and are what sentences stand for. Each of these has been argued to fall prey to the ‘slingshot.’ Another metaphysical view under attack has to do with notions such as necessity, possibility, causation, and temporal order. (Call these ‘MCT properties,’ for modal-causal-temporal.) At first glance, it seems that whether an object has an MCT property does not depend upon what the object is called. Call this its ‘first feature.’ Just as the rose smells sweet under any name, the rose is **necessarily** a flower no matter what it’s called; and its thorns **caused** this little cut on my hand, no matter what it’s called; and it bloomed **before** July 1, no matter what it’s called. MCT properties, that is, seem to be sensitive to the nature of the thing itself, not to how we speak about it. Even more obviously, just because one object has the property of being necessarily a flower, it doesn’t follow that every object does. And just because something caused that cut on my hand doesn’t mean that any old thing did. Similarly for temporal order: that the rose bloomed before July 1 doesn’t entail that anything you wish to pick happened before July 1. Call that its ‘second feature.’ Curiously, the slingshot argument has been used to show that, given an added complication about logical equivalence that will be explained below, these two supposed features of MCT properties can’t both apply. If we insist that not every truth is necessary, that not every event caused such-and-such, that not every event is temporally prior to so-and-so, then we must grant that whether something has an MCT property depends on what name is used for it. In this latter respect, MCT properties must be radically different from smelling sweet. (Terminological note: In what follows, I’ll speak of ‘operators’ when I mean ‘words that either modify one sentence, or connect two sentences.’ Thus all of ‘necessarily,’ ‘possibly,’ ‘because’ and ‘before’ are operators in my sense. Note also that the results to be discussed do not merely apply to words that modify and connect sentences: the arguments presented could easily be extended to expressions that combine sentences with subsentences, e.g., ‘[_S Juana died] before [_{NP} the American Civil War]’ combines a sentence with a noun phrase. To keep things simple, however, I will focus on operators combining or modifying sentences.)

Substitution *Salva Veritate*

I have quickly canvassed three targets of the slingshot: world-bound facts, the proposition as sentence meaning, and the claim that MCT properties have the two features introduced just above. Eventually I will explain how technical points about substitutivity *salva veritate* – the second element in the article’s

title – can be used to call facts *et al.* into question. But I need to start with what substitution *salva veritate* is, the varieties of substitutional contexts (i.e., for singular terms and for sentences), constraints on such substitution, etc.

I will begin with substitutivity of singular terms. Shakespeare famously said that a rose by any other name would smell as sweet. This certainly seems right: you don't change the smell of a rose just by renaming it. We can put his point about roses less eloquently as: 'You may substitute any coreferential name for "This rose" in "This rose smells sweet" and preserve truth.' Though wordy, this also seems right, and for just the same reason: a name change doesn't yield an odor change. Nor is '—— smells sweet' an isolated example. A dog by any other name would weigh just as much, would have the same number of hairs, would have the same size ears, etc. This is substitution *salva veritate*: 'substitution while preserving truth.'

Interestingly, not all contexts are like '—— smells sweet' or '—— weighs 28lb.' Sometimes when you change the label of an object, you don't preserve the truth of the whole sentence. Consider an example adapted from W. V. O. Quine:

- (1) Andre the Giant was so called because of his size
- (2) André Roussimoff was so called because of his size

Sentence (1) is true: the famous wrestler adopted that name precisely because he was so big. But sentence (2) is false. Surprisingly, this is the case even though 'Andre the Giant' and 'André Roussimoff' refer to the very same person. So, unlike roses and being sweet smelling, a wrestler by any other name would not automatically be so called because of his size; if he happens to have the property of being so called because of his size under all of his names, that would be the merest coincidence.

Of course what's special and different about '—— was so called because of his size' is that it explicitly makes reference to the name of the person being discussed: this curious predicate applies to linguistic things, i.e., names, not (just) to the person herself. That's why you can't put in any name you like for Andre in (1): because the sentence talks about words, specifically about names. We can see this point still more clearly with quotation marks. As it happens, the city of Toronto is also known as Hog Town. These two names refer to the same place. (Toronto used to be a major center for pork processing.) Yet (3) is true, while (4) is false:

- (3) 'Toronto' contains exactly three vowels
- (4) 'Hog Town' contains exactly three vowels

In this case it should be no surprise that you cannot substitute coreferring names in the context "—— contains exactly three vowels," because this predicate is obviously about the name, not about its referent. In contrast, '—— smells sweet' does not make reference to the linguistic item employed, but is instead wholly about the flower – that's why you can substitute whatever name you like. And '—— is so called because of his size' makes reference to both the denotation and the name: it's Andre the man who is being discussed, but 'so called' makes reference to his name as well. In sum, we have contexts which don't allow substitution of coreferring names *salva veritate* (both the kind which is just about words, as in quotation mark contexts, and the kind which is about words and nonwords, as in '—— was so called'), and contexts which do allow substitution *salva veritate* (the '—— smells sweet' and '—— weighs 28lb' kind).

The excursus into substitution principles continues. I've considered one kind of thing that can be substituted, namely singular terms. And I've considered one constraint on truth-preserving substitution, namely that the predicate apply to the thing denoted, not to the words that denote. Another kind of substitution deals not with names (such as 'Andre' and 'Toronto'), but rather with sentences. In some linguistic contexts, you can preserve truth by substituting sentences that have the same truth value. This works, for instance, with 'and,' 'or,' and the other truth functional connectives familiar from classical logic. To take one example, sentence (5) is true because both of its conjuncts are true. And since (6) is true as well, sentence (7) must be true, precisely because (7) results merely from substituting the true sentence (6) in for the second conjunct of (5).

- (5) Ottawa is the capital of Canada and Toronto is west of Ottawa
- (6) Vancouver is west of Edmonton
- (7) Ottawa is the capital of Canada and Vancouver is west of Edmonton

But many linguistic contexts don't preserve truth when one interchanges sentences that happen to have the same truth value. Take 'Just about everyone knows that ——.' Sentence (8), when embedded in this context, yields the true (9):

- (8) Two plus two is four
- (9) Just about everyone knows that two plus two is four

Now, sentences (6) and (8) have the same truth value: they are both true. Yet if we substitute (6) for (8), in 'Just about everyone knows that ——,' the result is

(10), which is not true. Canada's geography just isn't that widely known.

- (10) Just about everyone knows that Vancouver is west of Edmonton

So, unlike in the case of 'and,' we can go from the true (9) to the false (10) by substituting one true sentence for another. Thus, truth is **not** preserved under substitution, in the scope of 'Just about everyone knows that —.' This provides an example of not being able to substitute material equivalents *salva veritate*.

One last substitution principle. There are some contexts which allow substitution of logically equivalent sentences *salva veritate*. One example is 'entails.' Any two sentences which are logically equivalent entail the same things. Thus given the truth of (11) and the logical equivalence of (12) and (13), one can derive (14) by substitution.

- (11) That it's not the case that either Clinton is dead or Bush is dead entails that it's not the case that Clinton is dead
- (12) It's not the case that either Clinton is dead or Bush is dead
- (13) It's not the case that Clinton is dead and it's not the case that Bush is dead
- (14) That it's not the case that Clinton is dead and it's not the case that Bush is dead entails that it's not the case that Clinton is dead

But not all verbs are like this. Consider words like 'know,' 'believe,' 'expect' and so on – so-called propositional attitude verbs. Not only can one not automatically substitute true-for-true sentences, while preserving the truth of the whole, one cannot even substitute a **logically equivalent** sentence while guaranteeing truth. One way to see this is to consider that there are extremely complicated, and also very simple, ways of expressing claims which are logically equivalent. Put in the simple way, a child might know that the claim is true; but put in the extremely complex way, he might not. For instance, little Billie, a five-year-old, might expect that it will snow in January. That's simple enough for a child that age. But (15) is logically equivalent to the very complex (16):

- (15) It will snow in January
- (16) $\{x: x = x \ \& \ \text{It will snow in January}\} = \{x: x = x\}$

Now, Billie might well expect the former, yet not at all expect the latter.

Since logical equivalence of sentences of the form (15) and (16) will play a key role below, let me say rather more about it. Take ' $\{x: x$ is an odd number less than 27}' as an example. It refers to a set – specifically, a certain set of odd numbers. For what follows, it's

helpful to think of belonging to a set as involving meeting, or failing to meet, one or more conditions. For instance, an object belongs to $\{x: x$ is an odd number less than 27} if it is both an odd number, and less than 27. An object fails to belong to the set if it fails to meet either of these. Taking membership to involve meeting or failing to meet conditions, consider now the set $\{x: \text{Ottawa is the capital of Canada}\}$. At first glance this looks like an odd set, but the general rule still applies: an object belongs to this set if it meets the condition that Ottawa is the capital of Canada. Now, since every object meets that condition, every object belongs to that set. With this in mind, consider the first half of (16). This phrase stands for the set of x s such that x is self-identical and it will snow in January. So, there are two conditions that must be met by an object, in order for it to be in the set: the object must be self-identical, and it must be the case that it will snow in January. The first condition is trivially satisfied by every object, however. So, in a way parallel to $\{x: \text{Ottawa is the capital of Canada}\}$, whether an object gets into the set depends wholly upon whether it will snow in January: if it will snow in January, every object meets the two conditions for inclusion; if it will not snow in January, no object meets them. In this way, ' $\{x: x = x \ \& \ \text{It will snow in January}\}$ ' comes to denote the set of all individuals, if it will snow in January. Now, the set of all individuals is also, of course, what ' $\{x: x = x\}$ ' denotes. So the statement of their numerical identity, i.e., (16), is true if it will snow in January. On the other hand, if it won't snow in January, then ' $\{x: x = x \ \& \ \text{It will snow in January}\}$ ' denotes the null set: no object satisfies the two conditions for inclusion in this set, viz., that it be self-identical and that it will snow in January. Hence, if it won't snow in January, the statement of identity between ' $\{x: x = x \ \& \ \text{It will snow in January}\}$ ' and ' $\{x: x = x\}$ ' is false: the empty set, which is what the left-hand side of the equation would denote, does not equal the set of all objects, which is what the right-hand side denotes.

In short, the two sentences (16) and 'It will snow in January' are true and false together in every circumstance. So, they are logically equivalent. Nevertheless, Little Billie, I said, may well expect that it will snow in January; but, surely, it's not the case that little Billie expects that the set whose members are such that they are self-identical and it will snow in January is identical with the set whose members are self-identical! So, substitution of (16) for the logically equivalent (15), in the scope of 'expects that —,' does not necessarily preserve truth. If truth is preserved, it's because of a mere coincidence, namely that the person just happens to expect both things.

So, we have sentence operators like ‘and’ that allow substitution *salva veritate* of merely materially equivalent sentences – i.e., of sentences which just happen to have the same truth value. And we have sentence operators like ‘know’ and ‘expect’ that don’t allow that kind of substitution, and don’t even allow substitution of logically equivalent sentences. This completes my explanation of substitution *salva veritate* of sentences. We also surveyed substitution of singular terms. Many contexts allow this: ‘—— weighs 28 lb’ and ‘—— smells sweet.’ Some do not: ‘—— was so called because of his size.’ The next step is to put the notion of substitution *salva veritate* to work.

The Argument: The Slingshot Itself

Let’s take stock. In the first section we encountered three metaphysical views. The first two involved facts and factlike propositions: that facts are made up of objects, properties, and relations, that factlike propositions exist and serve as the meanings of natural language sentences. The third involved MCT properties having two features: of applying to objects themselves, and not applying to all truths.

Given the concepts introduced above, we can now rephrase this third metaphysical view, and add a third presumption about logical equivalence:

- i. Coreferential singular terms may be substituted one for another within the scope of MCT words without altering the truth of the whole sentence;
 - ii. Logically equivalent sentences may be substituted one for another within the scope of MCT words without altering the truth of the whole sentences.
- But
- iii. Sentences which are merely materially equivalent – i.e., which just happen to have the same truth value – may not be so substituted.

To introduce one last piece of terminology, this is to say that MCT words aren’t hyperintensional (i.e., they meet both (i) and (ii)) but they also aren’t transparent (i.e., they meet (iii)).

Before providing the language-based argument that none of these metaphysical views are really true, let’s quickly recall why it matters. What is attractive about the first two metaphysical views? Why worry about giving them up? First, as stressed above, a rose by any other name would smell as sweet. Second, we can’t have every linguistic change in the sentence giving rise to a different proposition. Third, and related to the second, we want the same proposition to give the meaning of sentences in different languages. And

why is it attractive to say that MCT words satisfy (i)–(iii)? Well, it would seem that ‘necessarily,’ ‘possibly,’ and other modal modifiers would allow substitution fairly freely, since, as noted, they don’t (seem to) apply to linguistic items. Like ‘—— smells sweet,’ these terms seem to be entirely about things, not at all about words. It’s Shakespeare’s rose point once again. For example, starting with modal operators, if Andre the Giant was necessarily a human being (and it seems that he was), then André Roussimoff was necessarily a human being too: it doesn’t matter what name we use for him, **he** is just as necessarily human. Similarly for causal operators: if Andre the Giant died because he had a heart attack, then André Roussimoff died because he had a heart attack. The phrase ‘—— died because he had a heart attack’ is about the person alone, regardless of how that person is denoted. Nor, turning to the temporal ‘before,’ could Andre the Giant die before Clinton was president, while André Roussimoff did not. As with ‘because,’ ‘before’ just doesn’t seem to work like ‘—— was so called because of his size’ and “——” contains exactly three vowels.’ It seems to work like ‘—— smells sweet’ and ‘—— weighs 28lb.’ It is this kind of reasoning which supports the claim that ‘necessary,’ ‘possibly,’ ‘before,’ and ‘because’ meet condition (i): coreferring singular terms may be substituted one for another, within the scope of **these** words, without altering the truth of the whole sentence. Moreover, unlike ‘know,’ ‘expect,’ and ‘believe,’ logically equivalent facts are either both necessary, or both not necessary, either both possible, or both not possible, either both before a certain event, or not. Thus, MCT words meet (ii) as well. So, these operators do allow substitution of the first two kinds – they are not hyperintensional. Yet, or so it would seem, you can’t substitute any old true sentence for ‘Andre the Giant is human’ in ‘It is necessary that Andre the Giant is human’: substitution of material equivalents is not sanctioned in the scope of modal modifiers. Nor can you substitute any old true sentence for ‘André Roussimoff died of a heart attack’ in ‘André Roussimoff died of a heart attack before Clinton was president’; nor in ‘Little Billie cried because André Roussimoff died of a heart attack.’ So MCT words meet (iii) too. (One might sum up by saying that MCT words are a mite promiscuous in terms of what substitutions they’ll allow, but it’s not a free-for-all either.)

Despite their attractiveness, however, and contrary to a once widely assumed semantics, if the slingshot argument works, MCT words either must not allow substitution of coreferring names or logically equivalent sentences, or they must allow substitution of sentences that just happen to be true together. And, despite the attractions of the metaphysical views,

there can be at most one world-bound fact, and it patently cannot serve as the meaning of all sentences!

The argument, at last. As the title of this article suggests, the ‘slingshot’ in question is not a Y-shaped frame spanned by an elastic; it is, instead, an argument. It was dubbed ‘the slingshot’ by Jon Barwise and John Perry, because it “is so small, seldom encompassing more than half a page, and employs such a minimum of ammunition” (Barwise and Perry, 1981: 398). Moreover, like the eponymous Y-shaped frame, it can, despite its modest make-up, be used to attack some very significant foes.

I will present the argument in two stages. First, I will give it in the abbreviated format one often encounters in journal articles and such. Second, I will offer a more extended version of the argument.

As Barwise and Perry (1981) and Neale (1995) both note, variations on the argument, in its very brief form, seem to have been first formulated independently by Alonzo Church (1943: 299–300, 1956: 24–25) and Kurt Gödel (1944: 128–129). But I will focus on the best-known variant of the argument, that due to Donald Davidson (1967a: 19). Since his presentation of the argument is especially condensed, and has left so many readers puzzled, I will cite it in full, and then try to explain it. (NB: I have altered Davidson’s numbering, and his logicomathematical notation, to bring it in line with what appears above.)

Davidson writes:

The difficulty follows upon making two reasonable assumptions: that logically equivalent singular terms have the same reference, and that a singular term does not change its reference if a contained singular term is replaced by another with the same reference. But now suppose that ‘R’ and ‘S’ abbreviate any two sentences alike in truth value. Then the following four sentences have the same reference:

- (a) R
- (b) $\{x: x = x \ \& \ R\} = \{x: x = x\}$
- (c) $\{x: x = x \ \& \ S\} = \{x: x = x\}$
- (d) S

For (a) and (b) are logically equivalent, as are (c) and (d), while (c) differs from (b) only in containing the singular term ‘ $\{x: x = x \ \& \ S\}$ ’ where (b) contains ‘ $\{x: x = x \ \& \ R\}$ ’ and these refer to the same thing if S and R are alike in truth value (Davidson, 1967a: 19).

As two generations of students can attest, this argument goes by very quickly. It is the ‘slingshot’ in its purest form. The first two sentences of the quotation essentially lay out, though in rather different terminology, conditions (i) and (ii) above. This is obscured by two things. First, because of the context in which he is writing, Davidson doesn’t explicitly limit his claims to sentences occurring within the scope of

words like ‘necessarily,’ ‘possibly,’ ‘before,’ and ‘because.’ (For a variant of the argument which is more explicitly restrictive along those lines, see Davidson, 1967b: 152–153, and also Davidson, 1969: 42.) Second, the relation between these first two sentences and conditions (i) and (ii) is obscured by the fact that Davidson is here assuming, for the sake of argument, that **sentences refer**, and hence just are a special kind of singular term; and that they refer specifically to **truth values**. Thus, when he says “a singular term does not change its reference if a contained singular term is replaced by another with the same reference,” this entails (i) as a special case: the special case where the containing ‘singular term’ is a sentence, and the contained singular term is a name. And when he says “logically equivalent singular terms have the same reference,” this yields, given his dictum about constant reference for the whole given constant reference of the parts, (ii): that logically equivalent sentences may be substituted, *salva veritate*, within larger sentences. Thus Davidson is here arguing that, despite appearances, any operator Φ , if it permits substitution of coreferential singular terms and substitution of logical equivalents within its scope, also permits substitution of sentences which are merely materially equivalent, i.e., which simply happen to share the same truth value. That is, using the terminology introduced above: if Φ is not hyperintensional, then it is transparent.

Let’s now unpack this. As I reconstruct it, the slingshot argument consists of two assumptions – which together essentially say that Φ is nonhyperintensional – plus three ‘lemmas’ based on logical relations. The assumptions are:

- A1. Substitution of **coreferential singular terms** in the scope of Φ will not change the truth value of the whole.
- A2. Substitution of **logically equivalent sentences** in the scope of Φ will not change the truth value of the whole.

The first two lemmas state logical equivalences:

L1. The sentence ‘ $\{x: x = x \ \& \ R\} = \{x: x = x\}$ ’ is logically equivalent to ‘R.’

Proof: If ‘R’ is true, then the left-hand side of the equation refers to the set of all individuals, because everything is such that it is self-identical and R obtains. And that’s what the right-hand side refers to as well. So, the equation as a whole is true, if ‘R’ itself is true. If ‘R’ is false, then the left-hand side of the equation refers to the empty set: if ‘R’ is false then nothing is such that it is self-identical and R obtains. But that’s not what the right-hand side refers to: the right-hand side still refers to the set of all individuals.

So, the equation as a whole is false, if ‘R’ is false. Thus the two sentences are logically equivalent.

L2. The sentence ‘ $\{x: x = x \ \& \ S\} = \{x: x = x\}$ ’ is logically equivalent to ‘S.’

Proof: Same as for L1.

The third lemma establishes a coreference relation:

L3. The singular term ‘ $\{x: x = x \ \& \ R\}$ ’ is coreferential with the singular term ‘ $\{x: x = x \ \& \ S\}$.’

Proof: Given that both sentences ‘S’ and ‘R’ are true, both of the singular terms in L3 refer to the set of all individuals. That is, everything is such that [it is self-identical and R]; and everything is such that [it is self-identical and S].

From the two assumptions about the nonhyperintensionality of Φ , and making use of the three lemmas about set theoretic relations, we will derive that Φ is transparent. The derivation, in effect, takes us to $\Phi(S)$ from $\Phi(R)$, for any two true sentences. This will show that mere material equivalence is, despite appearances, sufficient for substitution *salva veritate* within the scope of nonhyperintensional operators such as Φ . Thus, the general result is that if a context isn’t hyperintensional, it is transparent.

The derivation runs as follows. We start with (17):

(17) $\Phi(R)$

Given A2 and L1, we can substitute ‘ $\{x: x = x \ \& \ R\} = \{x: x = x\}$ ’ for the logically equivalent ‘R’ in (17) while preserving truth. This yields:

(18) $\Phi(\{x: x = x \ \& \ R\} = \{x: x = x\})$

Given A1 and L3, we don’t alter the truth value of (18) by substituting ‘ $\{x: x = x \ \& \ S\}$ ’ for the coreferential singular term ‘ $\{x: x = x \ \& \ R\}$ ’: these singular terms refer to the same set, i.e., the set of all objects, and we may thus substitute one name for the other. We may thus move from (18) to (19):

(19) $\Phi(\{x: x = x \ \& \ S\} = \{x: x = x\})$

Now, the final step in deriving $\Phi(S)$ from $\Phi(R)$ is the appeal to logical equivalence once again. Given A2 and L2, we can derive our desired target sentence (20) from (19):

(20) $\Phi(S)$

Being derivable in this way, (20)’s truth value must be the same as the original (17). The upshot is that if assumption 1 and assumption 2 are true with respect to a modifier/connective Φ (i.e., if Φ is not hyperintensional), then (20) may be derived from (17), for **any** true sentences ‘R’ and ‘S.’ Hence Φ , insofar as it’s not hyperintensional, does allow substitution of material equivalents after all: it is a transparent operator.

It’s worth stressing: Such a derivation can be constructed for **any** connective or modifier meeting conditions (i) and (ii): ‘necessarily,’ ‘possibly,’ ‘before,’ and ‘because,’ but also ‘the fact that — caused little Billie to cry’ and ‘the sentence “It will snow in January” stands for the proposition that —,’ and so on. So, the result isn’t merely about MCT properties – the third metaphysical view of the three surveyed in the first section – but includes the broader ones that there can be at most one world-bound fact, and that sentences cannot have factlike items as their denotations. Given the importance of these broader results, let us revisit the points in detail.

Aiming the Slingshot at Facts, and Factlike Things

Davidson and others have found the technical result important because they take it to show that sentences cannot stand for ‘facts,’ or factlike things. Or, more precisely, they cannot when facts are taken to be complexes of ordinary objects, properties, and relations. Indeed, they take it to show that such facts in the plural simply do not exist. The connection between the just-presented technical result and these broader theses is this: if there were multiple world-bound facts, and if sentences stood for them (or something like them, e.g., propositions), then expressions of the form (21) and (22) would have to meet conditions (i)–(iii). But we have just seen an argument that no expression can do this.

(21) The fact that Ottawa is the capital of Canada is different from the fact that —

(22) ‘Ottawa is the capital of Canada’ stands for the proposition that —

Here is the argument specifically against propositions, as I reconstruct it. (It should be straightforward to run the same sort of argument with respect to (21) and facts.) According to the view under consideration, which has facts/propositions containing ordinary objects and such, you don’t change from one proposition to another just by changing the name of the thing described. (As noted above, there are other ways of thinking about facts, and factlike things such as propositions, but the view that facts contain commonplace objects is the most intuitive for reasons already sketched.) Thus, changing the name only matters when the claim is actually about words, as in (1) and (3) above – and claims can’t always be about words; facts and propositions are to be autonomous from sentences. These provide good reasons why facts and propositions should be world bound, which in turn entails that the context in (22) must

allow substitution of coreferring singular terms. For quite similar reasons, it seems that two sentences which ‘say the same thing,’ but in different words, shouldn’t correspond to distinct facts/propositions; hence logically equivalent sentences should, it may seem, be substitutable in (22) without altering the truth of the whole. In this way we support variants on A1 and A2, here applied to the case (22). But once we have these two premises we can derive, using the argument above, that truth is retained if we substitute any true sentence.

Let’s walk through the derivation with an example, namely (23):

- (23) ‘Ottawa is the capital of Canada’ stands for the proposition that Ottawa is the capital of Canada

Suppose (23) is true, as it surely must be if sentences stand for factlike things. Given the logical equivalence of ‘ $\{x: x = x \ \& \ S\} = \{x: x = x\}$ ’ with S itself, for any sentence S , by A2 it follows that:

- (24) ‘Ottawa is the capital of Canada’ stands for the proposition that $\{x: x = x \ \& \text{Ottawa is the capital of Canada}\} = \{x: x = x\}$

Given that the singular term ‘ $\{x: x = x \ \& \ S\}$ ’ is coreferential with the singular term ‘ $\{x: x = x \ \& \ R\}$,’ as long as R and S are both true, we can apply this generalization to ‘Ottawa is the capital of Canada’ and ‘André Roussimoff died of a heart attack’ (both of which are true) to establish that:

- (25) ‘ $\{x: x = x \ \& \text{Ottawa is the capital of Canada}\}$ ’ is coreferential with the singular term ‘ $\{x: x = x \ \& \text{André Roussimoff died of a heart attack}\}$ ’

Now, we apply substitution of coreferring singular terms to (24). The result of substituting ‘ $\{x: x = x \ \& \text{André Roussimoff died of a heart attack}\}$ ’ for ‘ $\{x: x = x \ \& \text{Ottawa is the capital of Canada}\}$ ’ is:

- (26) ‘Ottawa is the capital of Canada’ stands for the proposition that $\{x: x = x \ \& \text{André Roussimoff died of a heart attack}\} = \{x: x = x\}$

The final step is to note that the complex identity statement about sets is logically equivalent to ‘André Roussimoff died of a heart attack.’ So, using substitution of logical equivalents a second time, now applied to (26), we can derive (27):

- (27) ‘Ottawa is the capital of Canada’ stands for the proposition that André Roussimoff died of a heart attack

In sum, given that facts and propositions are to be language independent in certain ways, “Ottawa is the capital of Canada” stands for the proposition that —’ should not be hyperintensional. But we

can use the slingshot to show that if the context in (22) is not hyperintensional, then it is transparent. That’s disastrous, however, because then this sentence ‘stands for’ any old true proposition! One can run precisely the same argument, beginning with (21), to show that there is only one world-bound fact.

Responses to the Slingshot Argument

It seems, very much contrary to initial appearances, that modal, causal, and temporal words lack at least one of these properties:

- i. Coreferential singular terms may be substituted one for another within the scope of these words without altering the truth of the whole sentence;
- ii. Logically equivalent sentences may be substituted one for another within the scope of these words without altering the truth of the whole sentences;
- iii. Sentences which are merely materially equivalent – i.e., which just happen to have the same truth value – may not be so substituted.

Also, because of substitution patterns like this, we seem also to be forced to give up world-bound facts and factlike propositions as the denotations of sentences. How is a metaphysician to respond?

One simply cannot give up (iii) as a condition on ‘the fact that —’ and such. That’s because, put in terms of an example, this would allow ‘the fact Ottawa is the capital of Canada’ to stand for the fact that Toronto is larger than Halifax! There are, however, at least three ways of saving MCT properties, facts, and factlike things such as propositions, in the face of the slingshot: two ways corresponding to (i) and one way corresponding to (ii). I will take them in reverse order.

First, one can reject the idea that logically equivalent sentences really do stand for a single fact, or factlike proposition. This is explicitly suggested by Barwise and Perry (1981). The proposal is that sentences with different subject matters do not stand for the same fact, even if they are true in precisely the same circumstances. Now (15) and (16), though logically equivalent, clearly have different subject matters, since only the latter talks about sets. So these sentences need not be taken to stand for the same fact. Granted, this makes facts, which are supposed to be built from ordinary elements of the world on the world-bound conception, look a bit more like the linguistic items that stand for them: facts get individuated in terms of meaning-related criteria such as ‘subject matter.’ However, this need not jeopardize the distinction between sentences and their meanings. For instance, we can still insist that sentences in different languages can stand for the same fact – as long as they have the same subject matter. And we can even

insist that sentences made up of coreferential words, put together in the same way, stand for the same fact. Maybe even transformed sentences can stand for the same fact, if the transformation does not alter the subject matter: e.g., ‘Moe kissed Juana’ might stand for the same fact as ‘Juana was kissed by Moe.’ One thus saves world-bound facts, and propositions, by rejecting (ii). (It’s less obvious that this move allows MCT properties to have the three features we want; but explaining the ins and outs of that would take us too far afield.)

Second, one can reply to the slingshot argument by denying that (i) really applies. This ‘it doesn’t apply’ move can be pursued in two different ways. Those who follow Frege (1892) maintain that you only have the same fact, or proposition, when not just the referent, but also the sense of the name, is the same. Famously, Frege introduced a two-level semantics, with guises-of-things occupying the level of sense, and the things themselves occupying the level of reference. Crucially for present purposes, being the same fact or proposition is supposed by Frege to require being the same at both levels. So, for example, ‘Elton John lives in England’ and ‘Reg Dwight lives in England’ don’t really correspond to the same proposition, because the two names differ in sense – even though these two names do refer to the same singer. This is to say that (i) ‘doesn’t apply’ in the sense that it’s false: coreferring singular terms can give rise to different facts/propositions. It should be clear that making this reply forces one to give up the world-bound conception of facts and factlike things. It is now not objects themselves, but objects described one way or another, that become the components of facts and propositions. This notion of proposition, with different propositions corresponding to the same referents arranged in the same way, is safe from the slingshot. As is the associated notion of fact. For this notion of fact/proposition allows one to reject substitution of coreferential singular terms in the contexts in question – and without that, the slingshot argument cannot get going.

Many contemporary philosophers find this Fregean approach unacceptable, however, for reasons already sketched: it seems like it’s the rose itself which smells sweet, not the rose under a guise; and it’s the man himself who died of a heart attack; and so on. So it seems, there should be no more to facts than perfectly ordinary objects, properties and relations. Not wanting to abandon the world-bound conception of facts, their alternative response to the slingshot argument is to insist that (i), though it is true, cannot be used as the proponent of the slingshot wishes to, because things like ‘ $\{x: x = x \ \& \ \text{André Roussimoff died of a heart attack}\}$ ’ aren’t really singular terms. It is in this

sense that condition (i) doesn’t apply. As Russell (1905) insisted, say these theorists, such expressions are complex definite descriptions whose role is quantificational, not referential: ‘the set of objects such that ...’ (which is what this mathematical expression actually means), is not namelike; it is instead more like the quantifiers ‘every set of objects such that ...’ and ‘some set of objects such that ...’. And condition (i) above, along with A1, which derives from (i), do not say anything about being able to substitute quantificational expressions *salva veritate*: A1 and (i) only apply to expressions whose function is to pick out an object, without describing it. Put in terms of examples, while it’s quite true that ‘Elton John is a singer’ and ‘Reg Dwight is a singer’ both stand for the same fact, since these involve two genuine names for the entertainer, it nevertheless is not the case that the description-involving sentences ‘The Queen of England lives in London’ and ‘Canada’s head of state lives in London’ stand for the same fact. Descriptions do not ‘merely pick out the object.’ Indeed, say these followers of Russell, what the sentence containing ‘the Queen of England’ really means is: There exists exactly one queen of England, and every queen of England lives in London. Crucially, therefore, the introduction of things like ‘ $\{x: x = x \ \& \ \text{André Roussimoff died of a heart attack}\}$ ’ takes us out of the realm of names (where [i] and A1 apply), and into the realm of descriptions (where they don’t). Thus, when one appeals to A1 *et al.* – to move from (18) to (19) in the general case, and from (24) to (26) in the specific case of “S” stands for the proposition that —’ – a fallacy is committed. (Neale, 1995, following Gödel, 1944, gave roughly this reply to the slingshot. Barwise and Perry, 1981 made related points.)

To summarize, then, the slingshot is an argument that deploys very modest machinery to take on several very important metaphysical views. The views include the two broad ones, that there are world-bound facts in the plural, and that what sentences stand for are factlike things, and a related more technical view about what kind of sentence-modifying and sentence-connective operators there can be. The key move in the argument is to start with an arbitrary sentence containing ‘R’ and one of the problematic operators (‘necessarily,’ ‘before,’ ‘because,’ ‘it’s a fact that —,’ etc.), to create a logically equivalent sentence that has ‘ $\{x: x = x \ \& \ R\} = \{x: x = x\}$ ’ in the scope of the suspect operator. This latter sentence *ipso facto* contains the singular term ‘ $\{x: x = x \ \& \ R\}$,’ which, assuming both ‘R’ and ‘S’ are true, is coreferential with ‘ $\{x: x = x \ \& \ S\} = \{x: x = x\}$.’ Hence, one can deploy the substitution of logical equivalents and the substitution of coreferring singular terms to derive another arbitrary materially equivalent sentence now containing not

'R' but 'S' in the scope of the operator. This shows that if the operator is not hyperintensional, then it's transparent. The three ways to block this argument are (a) to give up substitution *salva veritate* of logical equivalents in the scope of the operator, (b) to follow Frege, and give up substitution *salva veritate* of coreferential singular terms (and with it the world-bound conception of facts) and (c) to insist that substituting expressions of the form ' $\{x: x = x \ \& \ R\} = \{x: x = x\}$ ' is not, in fact, a matter of substituting singular terms, since such expressions are not referential at all, but are instead quantificational expressions.

Pulling back from the details, the slingshot illustrates how careful reflection on a linguistic phenomenon can contribute to debates in metaphysics. In particular, we have taken as our example the role of various principles of substitution *salva veritate* in attacks on facts, propositions, and views about the properties of modal, causal, and temporal properties. The possible replies to the slingshot highlight still further, I think, the importance of linguistic details when discussing metaphysics.

See also: Mood, Clause Types, and Illocutionary Force.

Bibliography

- Barwise J & Perry J (1981). 'Semantic innocence and uncompromising situations.' *Midwest Studies in Philosophy* 6, 387–403.
- Church A (1943). 'Review of Carnap's *Introduction to Semantics*.' *Philosophical Review* 52, 298–305.
- Church A (1956). *Introduction to Mathematical Logic*. Princeton: Princeton University Press.
- Davidson D (1967a). 'Truth and meaning.' *Synthese* 7, 304–323.
- Davidson D (1967b). 'Causal relations.' *Journal of Philosophy* 64, 691–703.
- Davidson D (1969). 'True to the Facts.' *Journal of Philosophy* 66, 748–764.
- Frege G (1892). 'On sense and reference.' *Zeitschrift für Philosophie und Philosophische Kritik* 100, 25–50.
- Gödel K (1944). 'Russell's mathematical logic.' In Schilpp P A (ed.) *The Philosophy of Bertrand Russell*. Chicago: Northwestern University Press. 125–153.
- Neale S (1995). 'The philosophical significance of Gödel's slingshot.' *Mind* 104, 761–825.
- Quine W V O (1960). *Word and object*. Cambridge, MA: MIT Press.
- Russell B (1905). 'On denoting.' *Mind* 14, 479–493.
- Strawson P F (1950). 'Truth.' *Proceedings of the Aristotelian Society* 24, 129–156.

Modal Logic

M Cresswell, The University of Auckland, Auckland, New Zealand

© 2006 Elsevier Ltd. All rights reserved.

In classical propositional logic, all the operators are truth-functional. That is to say, the truth or falsity of a complex formula depends only on the truth or falsity of its simpler propositional constituents. Take the monadic operator \sim . This is interpreted to mean 'it is not the case that', and its meaning is defined by Table 1, called a truth table, in which 1 represents 'true' and 0 represents 'false'. This operator is truth-functional because the truth value of the complex proposition it forms is determined by the truth value of the proposition it starts with. If α has the value 1, then $\sim\alpha$ has 0. If α is 0, $\sim\alpha$ is 1. (I use standard operators for propositional logic: \sim for negation, \vee for (inclusive) disjunction, \wedge for conjunction, \supset for

(material) implication and \equiv for (material) equivalence.) Modal logic is concerned with understanding propositions about what **must** or about what **might** be the case, and it is not difficult to see how we might have two propositions alike in truth value, both true, say, where one is true and could not possibly be false and the other is true but might easily have been false. For instance it must be that $2 + 2 = 4$, but, although it is true that I am writing this entry, it might easily not have been. Modal logic extends the well-formed formulae (wff) of classical logic by the addition of a one-place sentential operator L (or \Box) interpreted as meaning 'it is necessary that'. Using this operator, an operator M (or \Diamond) meaning 'it is possible that' may be defined as $\sim L\sim$. The notation L and M for the necessity and possibility operators dates from Feys (1950) (for L) and Becker (1930) (for M). (For a history of notation, see Hughes and Cresswell, 1968: 347–349). The use of \Box for the necessity operator is due to F. B. Fitch and first appears in Barcan (1946); the use of \Diamond for M dates from Lewis and Langford (1932). In fact, any one of L or M can be taken as primitive and the other defined in terms of it. Another operator that may be defined in terms of L and M is the operator Q (or ∇), where $Q\alpha$ means that α is

Table 1 Truth table

$\sim\alpha$
01
10

contingent, that is, neither necessary nor impossible. $Q\alpha$ may be defined as $\sim L\alpha \wedge M\alpha$. Many uses of *possible* in natural language suggest that what is possibly true is also possibly false, and those uses are better served by Q than by M .

The modern development of modal logic dates from 1912 when C. I. Lewis published the first of a series of articles and books (culminating in Lewis and Langford, 1932) in which he expressed dissatisfaction with the notion of material implication found in Whitehead and Russell's (1910) *Principia mathematica*. In the system of *Principia mathematica* – indeed in any standard system of propositional calculus (PC) – there are found the theorems:

- (1) $p \supset (q \supset p)$
- (2) $\sim p \supset (p \supset q)$

The sense of (1) is often expressed by saying that if a proposition is true, any proposition whatsoever implies it; the sense of (2) is often expressed by saying that if a proposition is false, it implies any proposition whatsoever. Lewis did not wish to reject these. He argued that they are “neither mysterious sayings, nor great discoveries, nor gross absurdities” (Lewis, 1912: 522), but merely reflect the truth-functional sense in which Whitehead and Russell were using the word *imply*. But he also maintained that there is a sense of *imply* in which when we say that p *implies* q we mean that ‘ q follows from p ’ and that in this sense of *imply* it is not the case that every true proposition is implied by any proposition whatsoever or that every false proposition implies any proposition whatsoever. A brief account of the early history of modal logic is found in Hughes and Cresswell (1968: pt. 3, with Lewis's contributions documented in Chaps. 12–13).

Until the early 1960s, modal logics were discussed almost exclusively as axiomatic systems without access to a notion of validity of the kind used, for example, in the truth-table method for determining the validity of wff of the classical propositional calculus. The semantical breakthrough came by using the idea that a necessary proposition is one that is true in all possible worlds. But whether another world counts as possible may be held to be relative to the point we are at. So an interpretation for a modal system would consist of a set W of possible worlds and a relation R of accessibility between them – where the worlds accessible from a given world are often described as the worlds that world can see. For any wff α and world w , $L\alpha$ will be true at w iff α itself is true at every w' such that wRw' . Put formally, a frame is a structure $\langle W, R \rangle$ in which W can be any class at all, although its members are often called ‘worlds’ or ‘indices,’ and R is a relation between them. A model can then be based on that frame by

adding a value assignment V . V assigns to each variable p a truth value in each world. If we read $V(p, w) = 1$ (0) as ‘ V assigns p the value true (false) in world w ’, then we may, for instance, define the truth table for \sim so that $V(\sim\alpha, w) = 1$ if $V(\alpha, w) = 0$, and 0 otherwise, and we may define the disjunction operator \vee so that $V(\alpha \vee \beta, w) = 1$ iff either $V(\alpha, w) = 1$ or $V(\beta, w) = 1$. $V(L\alpha, w) = 1$ iff $V(\alpha, w') = 1$ for every w' such that wRw' . We define validity on a frame by saying that a wff α is valid on a frame $\langle W, R \rangle$ iff, for every model $\langle W, R, V \rangle$ based on $\langle W, R \rangle$ and for every $w \in W$, $V(\alpha, w) = 1$. The ideas that underlie this account of validity appeared in the late 1950s and early 1960s in the works of Kanger (1957), Bayart (1958), Kripke (1959, 1963), Montague (1960), and Hintikka (1961). Anticipations can be found in Wajsberg (1933), McKinsey (1945), Carnap (1946, 1947), Meredith (1956), Thomas (1962), and other works. An algebraic description of this notion of validity is found in Jónsson and Tarski (1951), although the connection with modal logic was not made in that article.

An axiomatic basis for a logical system consists of a selected set of wff, known as axioms, together with a set of transformation rules, licensing various operations on the axioms and on wff obtained from the axioms by previous applications of the transformation rules. The wff obtained from the axioms in this way, together with the axioms themselves, are known as the theorems of the system. The system of modal logic whose theorems are precisely the wff valid on every frame is known as the system K. This name, which has now become standard, was given to the system in Lemmon and Scott (1977) in honor of Saul Kripke, from whose work the way of defining validity for modal logic is mainly derived. The word ‘frame’ in this sense seems to have been first used in print in Segerberg (1968), but the word was suggested to him by Dana Scott. The axioms of K consist of all valid wff of PC together with the modal wff K:

$$L(p \supset q) \supset (Lp \supset Lq)$$

and it has the following three transformation rules:

- US (The Rule of Uniform Substitution): The result of uniformly replacing any variable or variables p_1, \dots, p_n in a theorem by any wff β_1, \dots, β_n , respectively, is itself a theorem.
- MP (The Rule of Modus Ponens, sometimes also called the Rule of Detachment): If α and $\alpha \supset \beta$ are theorems, so is β .
- N (The Rule of Necessitation): If α is a theorem, so is $L\alpha$.

Other modal systems may be obtained by adding extra axioms to K. Each of these will be a proper

extension of K (i.e., it will contain not only all the theorems of K but other theorems as well). Modal systems that contain K (including K, itself) together with US, MP, and N are commonly known as normal modal systems.

A wff valid on every frame is called K-valid, and the theorems of the system K are precisely those wff that are K-valid. It is important to be clear that this is a substantive fact and not something that is true by definition. To be a theorem of K is to be derivable from the axioms of K by the transformation rules of K; to be K-valid is to be valid on every frame, and the fact that a wff is a theorem of K iff it is K-valid is something we have to **prove**, not something we can assume. Similar remarks hold for extensions of K because different systems of modal logic can represent different ways of restricting necessity. It can then happen that whether a principle of modal logic holds or not can depend on the properties of the accessibility relation. We may have an axiomatic modal system defined without any reference to an account of validity and a definition of validity formulated without any reference to theoremhood in a system, and yet the theorems of that system are precisely the wff that are valid by that definition. To show that there is a match of this kind between a system and a validity definition, we have to prove two things: (1) that every theorem of the system is valid by that definition and (2) that every wff valid by that definition is a theorem of the system. If the first holds, we say that the system is sound, and if the second holds we say that it is complete, in each case with respect to the validity-definition in question.

The wff $Lp \supset p$ is not K-valid because it is not valid on a frame in which there is a world that is not accessible from itself. (Put p false in this world, and true everywhere else.) We could, however, add it as an extra axiom to obtain a system stronger than K itself. The system obtained by adding $Lp \supset p$ as a single extra axiom to K is usually referred to as T, and it has had a long history in modal logic dating from Feys (1937). Feys's own name for the system is 't' (it was first called 'T' by Sobociński, 1953). Feys derived the system by dropping one of the axioms in a system devised by Gödel (1933), with whom the idea of axiomatizing modal logic by adding to PC originates. Sobociński showed that T is equivalent to the system M of von Wright (1951); for this reason 'M' was occasionally used as an alternative name for T. T is K + T.

T: $Lp \supset p$

Although T is not valid on every frame it is valid on all frames in which R is reflexive, that is, frames in which every world is accessible from itself – where

wRw for every $w \in W$. It can be proved that the system T is sound and complete with respect to such frames. In any system containing T, $L\alpha$ is equivalent to $\alpha \wedge \sim Q\alpha$, and so necessity may be defined in terms of contingency. Without T, this cannot usually be done. An examination of when necessity can be defined in terms of contingency in systems that do not contain T may be found in Cresswell (1988).

If we interpret L as expressing obligatoriness (moral necessity), we are unlikely to want to regard $Lp \supset p$ as valid because it will then mean that whatever ought to be the case is in fact the case. There is, however, a formula that, like $Lp \supset p$, is not a theorem of K but that, under the moral interpretation, it is plausible to regard as valid, and that is the wff $Lp \supset Mp$. If Lp means that it is obligatory that p , then Mp will mean that it is permissible that p (not obligatory that not- p), and so $Lp \supset Mp$ will mean that whatever is obligatory is at least permissible. This interpretation of L is known as a deontic interpretation, and for that reason $Lp \supset Mp$ is often called D, and the system obtained by adding it to K as an extra axiom is known as the system D; that is, D is defined as K + D.

D: $Lp \supset Mp$

D is sound and complete with respect to the class of frames in which R is serial – for every $w \in W$, there is some w' , perhaps w itself, as in T, but perhaps not, such that wRw' .

In the early days of modal logic, disputes centered around the question of whether a given principle of modal logic was correct or not. Often these disputes involved formulae in which one modal operator occurs within the scope of another – formulae such as $Lp \supset LLp$. Sequences such as LL are known as iterated modalities, and it is tempting to consider the possibility that all iterated modalities might be equivalent to uniterated ones or to a small number of iterated ones. Consider the following:

(3) $Mp \equiv LMp$

(4) $Lp \equiv MLp$

(5) $Mp \equiv MMp$

(6) $Lp \equiv LLp$

None of these is a theorem even of T. Some of them can be derived from others. We can obtain (5) and (6) by adding $Lp \supset LLp$ to T; this system is known as the system S4. We can obtain all of (3)–(6) by adding $Mp \supset LMp$ to T; this system is known as the system S5. The names 'S4' and 'S5' derive from Lewis and Langford (1932: 501), in which systems deductively equivalent to these are the fourth and fifth in a series of modal systems. These systems too can be studied semantically. Suppose that R is required to be

transitive; that is, suppose that, for any worlds w_1 , w_2 , and w_3 , if w_1Rw_2 and w_2Rw_3 then w_1Rw_3 . If so, then $Lp \supset LLp$ will be valid, but if nontransitive frames are permitted it need not be. S4 is sound and complete with respect to all reflexive and transitive frames. Suppose that in addition R is symmetrical, that is, if wRw' then $w'Rw$. S5 is sound and complete with respect to all reflexive, transitive, and symmetrical frames.

Not all normal modal logics are so well-behaved, however. There exist systems that can be proved to be incomplete in the sense that there is **no** class of frames such that their theorems are precisely the wff that are valid on every frame in the class. One simple example is the system $KH = K + H$.

H: $L(Lp \equiv p) \supset Lp$

Every frame for H also validates $Lp \supset LLp$, but $Lp \supset LLp$ is not a theorem of KH. The incompleteness of this system is proved in Boolos and Sambin (1985). A simplified proof appears in Hughes and Cresswell (1996: Chap. 9). Much current research in propositional modal logic is devoted to the systematic study of large families of logics to examine the conditions under which they have, or lack, properties such as these.

Modal logic can be given a temporal interpretation. The propositional logic of linear time in which L means 'it is and always will be that ...' is called S4.3; it is S4 together with **D1**.

D1: $L(Lp \supset q) \vee L(Lq \supset p)$

The logic of discrete time, in which each moment has a unique successor, S4.3.1, is S4.3 + **N1**.

N1: $L(Lp \supset Lp) \supset p \supset (MLp \supset Lp)$

This logic has a venerable history. In 1957, Arthur Prior conjectured that S4 was the logic of time with this structure, but later authors proved it to be S4.3.1. Prior's name for this was D (for Diodorus Cronos, not to be confused with D for the deontic system). The history is told in Prior (1967: Chap. 2). Another logic with a philosophical and mathematical interest is one variously called KW or G. It is K together with **W**.

W: $L(Lp \supset p) \supset Lp$

One of its interpretations is that it expresses the logic of provability – with L meaning, 'it is provable that'. From a purely modal point of view, it is sound and complete with respect to the class of all finite frames in which R is transitive and asymmetrical. The name W is from Segerberg (1971: 84); it is called G in Boolos (1979). For a more recent survey of the history of provability logic, see Boolos and Sambin (1990). The system dates at least from Löb (1966).

An important modal notion is that of entailment. By this we understand the converse of the relation of following logically from (when this is understood as a relation between propositions, not wff); that is, to say that a proposition, p , entails a proposition, q , is simply an alternative way of saying that q follows logically from p or that the inference from p to q is logically valid. In modal logic, we can say that p entails q iff $L(p \supset q)$. But then we are faced with the following valid wff.

(7) $L((p \wedge \sim p) \supset q)$

(8) $L((q \supset (p \vee \sim p))$

With this interpretation, (7) means that any proposition of the form $(p \wedge \sim p)$ entails any proposition whatever, and (8) means that any proposition whatever entails any proposition of the form $(p \vee \sim p)$. Although this may seem strange, those who wish to reject either (7) or (8) have to face the following argument. The following principles seem intuitively to be valid:

1. Any conjunction entails each of its conjuncts.
2. Any proposition, p , entails $(p \vee q)$, no matter what q may be.
3. The premises $(p \vee q)$ and $\sim p$ together entail the conclusion q (the principle of the disjunctive syllogism).
4. Whenever p entails q and q entails r , then p entails r (the principle of the transitivity of entailment).

Lewis showed long ago (Lewis and Langford, 1932: 250–251) that by using these principles we can always derive any arbitrary proposition, q , from any proposition of the form $(p \wedge \sim p)$, as shown in Table 2. This derivation shows that the price that has to be paid for denying that $(p \wedge \sim p)$ entails q is the abandonment of at least one of the principles 1–4. The most fully developed formal response to these paradoxes consists of abandoning principle 3, the principle of disjunctive syllogism. Logics that do this are called relevance logics. A survey of relevance logic is found in Dunn (1986) and Mares and Meyer (2001).

Table 2 Derivation of propositions using the four principles

Principle	Proposition
	(i) $p \wedge \sim p$
From (i), by A	(ii) p
From (ii), by B	(iii) $p \vee q$
From (i), by A	(iv) $\sim p$
From (iii) and (iv), by C	(v) q
So by D , $(p \wedge \sim p)$ entails q	

First-order predicate logic can also be extended by the addition of modal operators. (For more details of modal predicate logic see Hughes and Cresswell, 1996: pt. 3; Garson, 1984.) The most interesting consequences of such extensions are those that affect mixed principles, principles that relate quantifiers and modal operators and that cannot be stated at the level of modal propositional logic or nonmodal predicate logic. Thus, (9) is valid, but (10) is not.

$$(9) \exists xL\alpha \supset L\exists x\alpha$$

$$(10) L\exists x\alpha \supset \exists xL\alpha$$

(Even if a game must have a winner there need be no one who must win.) In some cases the principles of the extended system will depend on the propositional logic on which it is based. An example is a formula studied by Ruth Barcan Marcus, who first considered combining modal logic with first-order predicate logic and introduced the formula in Barcan (1946). The Barcan formula (BF) is:

$$\forall xL\alpha \supset L\forall x\alpha$$

which is provable in some modal systems but not in others. If both directions are assumed, so that we have $\forall xL\alpha \equiv L\forall x\alpha$, then this formula expresses the principle that the domain of individuals is held constant over all possible worlds. A temporal version of the Barcan formula might make its interpretation clearer. Let us assume that everyone now in existence will die before 2154. Then, if L means 'it will always be the case that' and α is the wff 'x dies before 2154', then $\forall xL\alpha$ is true. But it is not at all likely that it will always be the case that everyone will die before 2154 because people who do not now exist may well live beyond then, and so $L\forall x\alpha$ will be false. This assumes that a quantifier only ranges over the individuals that exist at the time or in the world at which the wff is being evaluated, but indicates why BF has been a matter of controversy.

It is possible to have complete predicate logics whose predicate extensions are incomplete. One such is S4.2, which is S4 + G1.

$$\text{G1: } MLp \supset LMp$$

S4.2 is characterized by frames that are reflexive, transitive, and convergent, in the sense that, if a world w_1 can see two worlds w_2 and w_3 then there is a world w_4 that both w_2 and w_3 can see. But the predicate extension of S4.2 + BF is not characterized by any class of frames. Perhaps more interesting is the case of the systems S4.3.1 and KW (or G) previously mentioned, the logics of discrete time and of provability. It is established in Cresswell (1997) that the predicate logic characterized by all frames for these

systems is not recursively axiomatizable because they, together with a whole family of systems containing N1 and $Lp \supset LLp$, when combined with first-order logic, enable the expression of second-order arithmetic.

When identity is added, even more questions arise. The usual axioms for identity easily allow the derivation of LI:

$$(x = y) \supset L(x = y)$$

but should we really say that all identities are necessary? Take:

$$(11) \text{ the composer of 'Threnody for Mrs S' = the composer of 'Salm'}$$

An important feature of (11) is that it uses the phrases *the composer of 'Threnody for Mrs S'* and *the composer of 'Salm'*. Such phrases are often called definite descriptions, and they pose problems even in nonmodal predicate logic. One of the first to see this was Bertrand Russell, whose celebrated example was

$$(12) \text{ the present king of France is bald}$$

Because there is no present king of France, it would seem that (12) is false. But then it would seem that (13) is true.

$$(13) \text{ the present king of France is not bald}$$

But (13) does not seem true either, for the same reason as (12). Even worse is sentence (14).

$$(14) \text{ the present king of France does not exist}$$

Russell claimed that the phrase *the present king of France* does not function like a name at all, and his account of the matter can help with (11). If we follow Russell, (11) makes five claims:

1. At least one person composed 'Threnody for Mrs S.'
2. At most one person composed 'Threnody for Mrs S.'
3. At least one person composed 'Salm.'
4. At most one person composed 'Salm.'
5. $\forall x\forall y((x \text{ composed 'Threnody for Mrs S' } \wedge y \text{ composed 'Salm'}) \supset x = y)$.

Look carefully at claim 5. This claim is true but not necessarily true, so putting L in front of it gives us a false sentence. But LI does not license us to put L in front of claim 5. What LI does allow is for us to move from claim 5 to claim 6.

6. $\forall x\forall y((x \text{ composed 'Threnody for Mrs S' } \wedge y \text{ composed 'Salm'}) \supset L(x = y))$.

And it is less clear that claim 6 is at all objectionable.

Suppose, nevertheless, that we still wish to abandon LI. To falsify LI, we let the values of the variables be strings of objects that may coincide in some worlds but not in others. In the present example, letting x mean 'The composer of "Threnody for Mrs S"' would mean requiring its value in a world w be whoever in w composed 'Threnody for Mrs S,' while y could stand for whoever composed 'Salm' in w . But there is a problem. Allowing all such strings would make the wff (10) valid, and it was remarked that (10) could be false in a game in which it is necessary that some player will win, although there is no individual player who is bound to win. There is, however, one way in which we could make (10) sound plausible. Consider, for example, the expression *the governor-general of New Zealand*, as it may occur in a constitutional context. The law may specify that at a certain point the signature of the governor-general is required before an act of parliament becomes law; yet on one occasion the governor-general may be Michael Hardie-Boys and on another it may be Sylvia Cartwright. Thus, the phrase *the governor-general of New Zealand* does not designate any particular individual (as an assemblage of flesh and blood); yet we can in a sense think of it as standing for a single object, contrasted with *the prime minister* and so forth.

Such objects are often called intensional objects or individual concepts. In a logic in which the individual variables range over all intensional objects, (10) would be valid because, if it must be the case that there is someone who must sign acts into law, then although no individual person must be that someone, yet there is someone (viz. the governor-general) whose signature is required. An adequate semantics for the contingent identity systems that do not validate (10) would therefore have to place restrictions on allowable strings of objects and neither require that only strings consisting of the same member of D in each world should count as objects—for that would validate LI – nor that any string whatever of members of D should count as an object – for that would validate (10).

All the systems of modal logic considered so far have assumed a single necessity operator, with its possibility dual. An important class of systems with more than one is the class of tense logics. (For a survey see Prior, 1967.) A tense logic has two operators, L_1 and L_2 , where L_1 means 'it always will be the case that' and L_2 means 'it always has been the case that'. In frames for a tense logic, the associated relations R_1 and R_2 are so related that one is the converse of the other, that is, wR_1w' iff $w'R_2w$. Another way of introducing families of modal operators is suggested by a possible interpretation of modal logic in computer science. In this interpretation, the worlds are states

in the running of a program. If π is a computer program, then $[\pi]\alpha$ means that, after program π has been run, α will be true. If w is any world, then $wR_\pi w'$ means that state w' results from the running of program π . This interpretation of modal logic is called dynamic logic. What gives dynamic logic its interest is the possibility of combining simple programs to get more complex ones. Thus, if π_1 and π_2 are two programs, then the expression $\pi_1;\pi_2$ refers to the program 'first do π_1 and then do π_2 ', and $[\pi_1;\pi_2]\alpha$ means that α will be true if this is done. The relation corresponding to $[\pi_1;\pi_2]$ may be defined to hold between w and w' iff $\exists u(wR_{\pi_1}u \wedge uR_{\pi_2}w')$. Other computing operations can generate similar modal operators, with appropriate conditions on their accessibility relations (see Goldblatt, 1987: Chap. 10). It is also possible to develop dynamic predicate logic. For an introductory survey, see Goldblatt (1987: pt. 3).

The most general kind of possible-worlds semantics for propositional operators is based on that idea of the truth set of a formula. In any model, we can define $|x|$ as $\{w \in W: V(x,w) = 1\}$. In evaluating Lx in a world w , all the input that we require is to know which set of worlds forms the truth set of x . Whatever L means, what it has to do is to declare Lx true at w for some truth sets and false for others. So the meaning of L must specify which sets of worlds form acceptable truth sets in world w . These sets of worlds are called the neighborhoods of w , and a neighborhood frame for a language of modal propositional logic is a pair $\langle W, R \rangle$ in which W is a set (of worlds) and R is a neighborhood relation. (For some remarks on the history of neighborhood semantics, see Segerberg, 1971: 72.) A neighborhood relation is a relation between a world w and a subset A of W , and A is a neighborhood of w iff wRA . We then say that $V(Lx,w) = 1$ iff $wR|x|$. A frame of the kind assumed earlier, in which R is a relation between worlds, is often called a relational frame, and it is not difficult to see that every relational frame is a special case of a neighborhood frame. To be precise, a relational frame is a neighborhood frame in which for every $w \in W$ there is a set B of those and only those worlds that are accessible to w (i.e., B is the set of worlds w can 'see') and wRA iff $B \subseteq A$. What this means is that x 's truth set is a neighborhood of w iff it contains all the worlds accessible from w , which is of course precisely what the truth of Lx in a relational frame amounts to.

Neighborhood semantics can be devised for systems with operators taking more than one argument. A philosophically important example here is the logic of counterfactuals as developed in the late 1960s and early 1970s. We present here a version of the semantics found in Lewis (1973), but the same idea is also

found in Stalnaker (1968) and Åqvist (1973). Counterfactual logic is based on a dyadic operator $\Box \rightarrow$, where $\alpha \Box \rightarrow \beta$ means that ‘if α were to be the case then β would be the case’. Lewis’s idea is that, given a possible world w , some worlds are closer to w than others. If we write $w' <_w w''$ to mean that ‘ w' is closer to w than w'' is’, then the semantics for $\Box \rightarrow$ will be that $\alpha \Box \rightarrow \beta$ is true in w iff either α is not true at any world or there is a world w' at which α and β are both true that is closer to w than any world w'' at which α is true but β is not. A counterfactual frame can be described as a neighborhood frame in the following way. Because $\Box \rightarrow$ is dyadic, its neighborhood relation R will relate worlds to pairs $\langle A, B \rangle$ where $A \subseteq W$ and $B \subseteq W$. The standard rule for a dyadic operator δ using an associated neighborhood relation R is (15).

$$(15) \quad V(\delta\alpha\beta, w) = 1 \text{ iff } wR\langle\alpha, \beta\rangle$$

A frame $\langle W, R \rangle$ will be a counterfactual frame iff it is based on a nearness relation $<$ in such a way that $wR\langle A, B \rangle$ iff either (16) or (17).

$$(16) \quad A = \emptyset$$

$$(17) \quad \text{There is some } w' \text{ such that } w' \in A \cap B \text{ and for every } w'', \text{ if } w'' \in A \cap \neg B \text{ then } w' <_w w''$$

Which counterfactual logic we get will depend on what kind of conditions we put on $<$. For instance, under the plausible assumption that the closest world to w is w itself, we get a frame that validates the wff in (18).

$$(18) \quad p \supset ((p \Box \rightarrow q) \equiv q)$$

By contrast, on any plausible account of nearness, many wff that are valid for \supset fail for $\Box \rightarrow$. For instance in standard systems of counterfactual logic, neither (19) nor (20) is valid.

$$(19) \quad ((p \Box \rightarrow q) \wedge (q \Box \rightarrow r)) \supset (p \Box \rightarrow r)$$

$$(20) \quad (p \Box \rightarrow q) \supset (\sim q \Box \rightarrow \sim p)$$

Questions such as these bring us to the boundary between modal logic, metaphysics, and semantics and remind us of the rich potential that the theory of possible worlds has for illuminating such issues. The possible-worlds semantics can be generalized to deal with any operators whose meanings are operations on propositions as sets of possible worlds and form a congenial tool for those who think that the meaning of a sentence is its truth conditions and that these be taken literally as a set of possible worlds – the worlds in which the sentence is true.

Further information on modal logic, including more detailed bibliographical references, may be found in a number of works. See, for instance, Blackburn *et al.*

(2001), Bull and Segerberg (1984), Chargrov and Zaharyashev (1997), Chellas (1980), and Hughes and Cresswell (1996).

See also: Conditionals; *De Dicto* versus *De Re*; Fictional Discourse: Philosophical Aspects; Indexicality: Philosophical Aspects; Quantifiers: Semantics; Semantic Value; Tense and Time: Philosophical Aspects.

Bibliography

- Åqvist L (1973). ‘Modal logic with subjunctive conditionals and dispositional predicates.’ *Journal of Philosophical Logic* 2, 1–76.
- Barcan (Marcus) R C (1946). ‘A functional calculus of first order based on strict implication.’ *Journal of Symbolic Logic* 11, 1–16.
- Bayart A (1958). ‘La correction de la logique modale du premier et second ordre S5.’ *Logique et Analyse* 1, 28–44.
- Becker O (1930). ‘Zur Logik der Modalitäten.’ *Jahrbuch für Philosophie und Phänomenologische Forschung* 11, 497–548.
- Blackburn P, de Rijke M & Venema Y (2001). *Modal logic*. Cambridge, UK: Cambridge University Press.
- Boolos G (1979). *The unprovability of consistency*. Cambridge, UK: Cambridge University Press.
- Boolos G & Sambin G (1985). ‘An incomplete system of modal logic.’ *Journal of Philosophical Logic* 14, 351–358.
- Boolos G & Sambin G (1990). ‘Provability: the emergence of a mathematical modality.’ *Studia Logica* 50, 1–23.
- Bull R A & Segerberg K (1984). ‘Basic modal logic.’ In Gabbay D M & Guenther F (eds.) *Handbook of philosophical logic*, vol. 2. Dordrecht: Reidel. 1–88.
- Carnap R (1946). ‘Modalities and quantification.’ *The Journal of Symbolic Logic* 11, 33–64.
- Carnap R (1947). *Meaning and necessity*. Chicago: University of Chicago Press.
- Chargrov A & Zaharyashev M (1997). *Modal logic*. Oxford: Clarendon Press.
- Chellas B F (1980). *Modal logic: an introduction*. Cambridge, UK: Cambridge University Press.
- Cresswell M J (1987). ‘Magari’s theorem via the recession frame.’ *Journal of Philosophical Logic* 16, 13–15.
- Cresswell M J (1988). ‘Necessity and contingency.’ *Studia Logica* 47, 146–149.
- Cresswell M J (1997). ‘Some incomplete modal predicate logics.’ *Logique et Analyse* 160, 321–334.
- Dummett M A E & Lemmon E J (1959). ‘Modal logics between S4 and S5.’ *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 5, 250–264.
- Dunn J M (1986). ‘Relevance logic and entailment.’ In Gabbay D M & Guenther F (eds.) *Handbook of philosophical logic*, vol. 3. Dordrecht: Reidel. 117–224.
- Fey R (1937). ‘Les logiques nouvelles des modalités.’ *Revue Neoscholastique de Philosophie* 40, 517–553; 41, 217–252.

- Feys R (1950). 'Les systèmes formalisés Aristotéliennes.' *Revue Philosophique de Louvain* 48, 478–509.
- Garson J W (1984). 'Quantification in modal logic.' In Gabbay D M & Guentner F (eds.) *Handbook of philosophical logic*, vol. 2. Dordrecht: Reidel. 249–307.
- Gödel K (1933). 'Eine Interpretation des intuitionistischen Aussagenkalküls.' *Ergebnisse eines mathematischen Kolloquiums* 4, 34–40.
- Goldblatt R I (1987). *Logics of time and computation*. Stanford, CA: CSLI.
- Hintikka K J J (1961). 'Modality and quantification.' *Theoria* 27, 110–128.
- Hughes G E & Cresswell M J (1968). *An introduction to modal logic*. London: Methuen.
- Hughes G E & Cresswell M J (1996). *A new introduction to modal logic*. London: Routledge.
- Jónsson B & Tarski A (1951). 'Boolean algebras with operators.' *American Journal of Mathematics* 73, 891–939.
- Kanger S (1957). *Provability in logic*. Stockholm: Almqvist & Wiksell.
- Kripke S A (1959). 'A completeness theorem in modal logic.' *Journal of Symbolic Logic* 24, 1–14.
- Kripke S A (1963). 'Semantical analysis of modal logic I: normal propositional calculi.' *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 9, 67–96.
- Lemmon E J & Scott D S (1977). *The 'Lemmon notes': an introduction to modal logic*. Segerberg K (ed.). Oxford: Basil Blackwell.
- Lewis C I (1972). 'Implications and the algebra of logic.' *Mind* N.S. 21, 522–531.
- Lewis C I & Langford C H (1932). *Symbolic logic*. New York: Dover.
- Lewis D K (1973). *Counterfactuals*. Oxford: Basil Blackwell.
- Löb M H (1966). 'Extensional interpretations of modal logics.' *Journal of Symbolic Logic* 31, 23–45.
- Mares E D & Meyer R K (2001). 'Relevant logics.' In Goble L F (ed.) *The Blackwell guide to philosophical logic*. Oxford and Malden, MA: Blackwell. 280–308.
- McKinsey J C C (1945). 'On the syntactical construction of systems of modal logic.' *Journal of Symbolic Logic* 10, 83–96.
- Meredith C A (1956). *Interpretations of different modal logics in the 'property calculus'* Prior A N (recorded and expanded). Christchurch, NZ: Philosophy Department, Canterbury University College.
- Montague R M (1960). 'Logical necessity, physical necessity, ethics and quantifiers.' *Inquiry* 4, 259–269. (Reprinted in Montague 1974, 71–83.)
- Montague R M (1967). *Past, present and future*. Oxford: Oxford University Press.
- Montague R M (1971). *An essay in classical modal logic* (3 vols). Uppsala: Filosofiska studier.
- Montague R M (1974). *Formal philosophy*. New Haven, CT: Yale University Press.
- Prior A N (1957). *Time and modality*. Oxford: Oxford University Press.
- Prior A N (1967). *Past, Present and Future*. Oxford: Oxford University Press.
- Segerberg K (1968). *Results in non-classical logic*. Lund: Berlingska Boktryckeriet.
- Segerberg K (1971). *An Essay in Classical Modal Logic* (3 vols). Uppsala: Filosofiska studier.
- Sobociński B (1953). 'Note on a modal system of Feys-von Wright.' *Journal of Computing Systems* 1, 171–178.
- Stalnaker R C (1968). 'A theory of conditionals.' In Rescher N (ed.) *Studies in logical theory*. Oxford: Basil Blackwell. 98–112.
- Thomas I (1962). 'Solutions of five modal problems of Sobociński.' *Notre Dame Journal of Formal Logic* 3, 199–200.
- Wajsberg M (1933). 'Ein erweiterter Klassenkalkül.' *Monatshefte für Mathematik und Physik* 40, 113–126.
- Whitehead A N & Russell B A W (1910). *Principia mathematica* (3 vols). Cambridge, UK: Cambridge University Press.
- Wright G H von (1951). *An essay in modal logic*. Amsterdam: North Holland.

Modern Linguistics: 1800 to the Present Day

R Harris, Oxford, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

It is during the course of the 19th century that we first find linguists self-consciously making statements about the history of their own discipline. The very fact that they do so is part of the *prima facie* evidence for the existence of that discipline (as indeed it was intended to be). For, it need hardly be said, their main concern was not so much to establish what had happened in the past as to establish the contemporary

autonomy of their own branch of inquiry. For this purpose they usually adopted a familiar “torch of learning” model, in which knowledge is passed on from one age to the next. As successive generations fan the sacred Flame with their own contribution, it burns ever brighter. The current torch-holder often turns out to be the scholar telling the story, or a close colleague.

This pattern of historical self-justification has been repeated more or less without interruption in linguistics from 1800 down to the present day, beginning with the Comparative Philologists and continuing with the Neogrammarians. A classic case

from the early 20th century is the introductory chapter in F. de Saussure's *Cours de linguistique générale* (1916), which itself recruits the Comparative Philologists and the Neogrammarians as honorable predecessors. Another is chapters 1–4 of *Language: its nature, development and origin* (Jespersen, 1922). A third is chapter 1 of *Language* (Bloomfield, 1933). A fourth and even more blatant example (if that were possible) is *Linguistics* (Fries, 1966), which leads up to what Fries calls triumphantly “the breakthrough” of American structural linguistics and the arrival of transformational-generative grammar. Endemic to this self-justificatory pattern is the notion that one's predecessors, although great scholars, never quite got it right (or even, occasionally, got it wrong, thus all but extinguishing the Flame).

Underlying most of these self-justificatory historical accounts lies a concern to show that linguistics is – or at last has now, i.e., in the author's own generation, become – a ‘science.’ This anxiety is bound up with the whole question of the academic prestige of ‘the sciences’ as developed in 19th- and 20th-century universities in Europe, and the diminishing prestige of ‘the arts.’ ‘Science’ was from the beginning a watchword or slogan that modern linguistics appropriated as its own, thus implicitly distinguishing the subject from all ‘pre-scientific’ studies of language and languages. It is emphatically reaffirmed in the glossary to the first edition of this encyclopedia, which contains the blunt propagandist equation “linguistic science = linguistics” (Asher, 1994).

The brief account that follows describes the main phases usually identified in this self-serving tradition of historiography. It makes no attempt to list all the individual ‘schools’ and ‘movements’ that have contributed to modern linguistics, let alone all the eminent scholars. Nor does it cover ‘applied’ linguistics in such diverse areas as language acquisition, language pathology, language planning, language teaching, and psycholinguistics.

Comparative Philology

Modern linguistics is usually portrayed as arising out of a preceding inquiry called Comparative Philology or Comparative Grammar. The Comparativists were already anxious to distinguish themselves from an earlier scholarly tradition that they often referred to as Classical Philology. Adopting a “torch of learning” model requires an identification of some person or persons who first lit the torch. That role was retrospectively thrust, by the Comparativists and their successors, upon Sir William Jones (1746–1794).

According to later scholars seeking to validate their version of the history of linguistics, the stimulus that

Jones provided to 19th-century developments was ‘immense’ (Robins, 1994). But this overenthusiastic claim is based mainly on the endless quotation of one paragraph from a single lecture that Jones gave in India to the Asiatic Society in 1786. There he proposed that Sanskrit, Greek, and Latin were languages that might be derived from “some common source, which, perhaps, no longer exists.” He also suggested in the same paragraph that Gothic, Celtic, and Old Persian might have belonged to the same family.

On this flimsy evidence, Jones could hardly have anticipated that later generations would hail him as a ‘founder’ of Comparative Philology, much less as ‘the first modern linguist.’ Jones was a newly arrived British judge sent out from London to administer colonial justice in India. What he saw was the practical advantage of having at his disposal a digest of Hindu and Islamic law. For this purpose, a knowledge of Sanskrit was essential, and he applied himself to it. The irony is that Jones himself had no great regard for philological studies as such. As far as he was concerned, languages were “mere instruments of real learning.” They were not objects of study in their own right.

But that, precisely, was the central notion of the Comparativists. In their view, languages were in themselves organic growths, developing in accordance with their own organic nature, and therefore qualifying as raw ‘materials’ for a modern science of language. On this basis, as F. M. Müller (1823–1900), first professor of Comparative Philology at Oxford, insisted, it was possible to envisage the study of languages as comparable to such natural sciences as biology and astronomy. But no such project was further from the mind of Jones.

The scholars who, in practice, set up Comparative Philology as an academic discipline were J. Grimm (1785–1863), R. Rask (1787–1832), and F. Bopp (1791–1867). Their work had as its focus the comparative study of the Indo-European languages. They were not engaged in what later came to be known as ‘general linguistics,’ except insofar as their methods and arguments presupposed certain assumptions about the ‘nature’ of language. One of their preoccupations was the historical task of ‘reconstructing’ the unattested ancestral language from which, they assumed, all Indo-European languages had sprung. In this respect, they placed previously inchoate discussions of linguistic origins on a new footing.

The Neogrammarians

The Neogrammarians are seen as marking an ‘advance’ on Comparative Philology, inasmuch as

they endeavored, in the mid 19th century, to establish 'laws' of linguistic development. Discovering 'laws' was then regarded as a *sine qua non* for setting up a 'science.' The principal figures in the Neogrammarian (*Junggrammatiker*) movement were acknowledged to be H. Osthoff (1847–1909), K. Brugmann (1849–1919), A. Leskien (1840–1916), and H. Paul (1846–1921). Their main claim to fame was their formulation of the principle that the 'laws' of sound change hold without exception. Thus if, for example, an intervocalic stop consonant becomes a fricative in one word, this is predicted to happen in all such words when that sound occurs in a similar environment in the same language. In this respect, the Neogrammarians supposedly brought linguistics in line with the 19th-century conception of the natural sciences. Osthoff spoke of the "blind necessity" with which sound laws operate. (Cf. the Newtonian law of gravitation.) The claim itself is virtually worthless, since any instance of an exception could always, for the Neogrammarians, be accommodated by *ad hoc* modification of the 'law' in question. We end up with mere statements of what has been observed to happen with regularity in the past. The exceptionless laws are no more than products of their own hind-sighted formulation.

Saussurean Structuralism

Although F. de Saussure (1857–1913) paid great respect to the Neogrammarians, his approach opened up a quite different chapter in academic language studies. The posthumously published *Cours de linguistique générale* (1916), based on lecture notes taken by his students at the University of Geneva, came to be regarded as the Magna Carta of modern linguistics. In this work, linguistics is presented as one branch of a more all-embracing study of signs, which Saussure referred to as 'semiology' (*sémiologie*).

Saussure did not believe in linguistic 'laws.' He saw 19th-century 'historical' linguistics as a muddle and thought that the Neogrammarians offered no insight at all into what a language was for its speakers. He accordingly drew a fundamental distinction between 'synchronic' and 'diachronic' linguistics. Synchrony is often equated, but mistakenly, with a study of the present-day language. Nor is Saussure's *linguistique diachronique* a synonym for *linguistique historique*, as will be evident to anyone who reads attentively the section on that subject in the *Cours de linguistique générale*. But it was often misunderstood by later linguistic theorists.

Saussure did not mount an all-out attack on comparative and historical linguistics, but he shunted those inquiries into an intellectual siding from which

they could make no effective contribution to answering the main questions that linguistic theory had to tackle.

For Saussure, *linguistique diachronique* was an academic edifice set up by the linguist, with a perspective altogether different from that available to the individual member of a linguistic community at any one time. *Linguistique synchronique*, on the other hand, was an attempt to capture the psychological reality underlying the way a typical member of the linguistic community uses its communal language. This distinction was, and always will be, controversial. Saussure himself did not help matters by sometimes appearing to identify the synchronic viewpoint with that of the traditional (prescriptive) grammarian.

Saussure's distinction between *langue* (the collective language system) and *parole* (the linguistic act of the individual) also remained a bone of contention. Although Saussure neither coined nor used the term "structuralism," this was the label later attached to his conception of synchronic analysis. Its principal feature is that forms and meanings are not to be identified by matching them with anything external to language (e.g., sounds or physical objects), but solely by reference to internal (structural) contrasts arising within the language itself. Thus, for example, the form of the English word /bat/ cannot be stated in terms of the sequence of sounds [bat] but only by listing the contrasts that distinguish this word from /kat/, /fat/, /hat/, etc. This, for Saussure, is a necessary consequence of realizing that different languages operate with different sets of phonological contrasts: there is no fixed and finite set of phonemes valid for all languages. Similarly, the meaning of a word like *horse* cannot be stated by identifying a certain species of animal and saying that *horse* is the name of this species. Why not? Because different languages recognize different biological species. Thus Saussure's view of meaning departed radically both from traditional lexicography and from that of the founder of modern semantics, his older contemporary M. Bréal (1832–1915). Saussure rejects the entire nomenclaturist approach to language entrenched in the Western tradition. It follows from this that all languages are irreducibly different. They are not just alternative ways of expressing some universal set of concepts, as Aristotle and many others had supposed.

By adopting this uncompromisingly holistic approach to linguistic analysis, Saussure in effect pulled the theoretical rug from under both Comparative Philology and historical linguistics as practiced by the Neogrammarians. His ideas were taken up and developed, particularly in Europe, by the so-called Prague school of linguistics. Two of its leading members were

N. Trubetzkoy (1890–1938), whose work on phonology became highly influential, and R. Jakobson (1896–1983). Another branch of structural linguistics emerged in Denmark as ‘glossematics,’ whose leading exponent was the Danish linguist L. Hjelmslev (1899–1965). In France, G. Guillaume (1883–1960) was responsible for a highly original exploration of – and reaction against – some Saussurean ideas, in propounding what he called the ‘psychomechanics’ of language.

During Saussure’s lifetime, two other approaches – linguistic geography and linguistic anthropology – came to prominence in language studies.

Linguistic Geography

In Europe, dialectologists began the task of documenting current linguistic usage and the geographical distribution of linguistic features. This led to the compilation of the first linguistic ‘atlases.’ The earliest (1881) was that of the German dialectologist G. Wenker, who sent questionnaires to 30,000 German schoolteachers in an attempt to survey features of local pronunciation. A Romanian linguistic atlas was published by G. Weigand in 1909, compiled on the basis of direct interviews with informants. In France, J. Gilliéron and E. Edmont published (1902–1910) a linguistic atlas based on interviews carried out at more than 600 localities. All this early work was done without the benefit of sound-recording apparatus and thus depended very much on the linguist’s ear and the transcription system used. From it emerged, however, one new (and controversial) theoretical concept: the ‘isogloss.’ Isoglosses were lines drawn on a map, demarcating areas in which a particular feature occurred and separating those from areas of non-occurrence. The study of isoglosses in turn gave rise to controversial conclusions concerning the existence of ‘dialects,’ which, according to some investigators, proved to be illusory.

J. Gilliéron (1854–1926), on the basis of his cartographical studies, pronounced the ‘bankruptcy’ of traditional etymology and reached the famous conclusion that “each word has its own history.” When this conclusion is combined with doubt about the existence of dialects, it leads directly to doubt about the existences of ‘languages’ too.

Linguistic Anthropology

In America during the same period, anthropologists provided a significant input to language studies, in particular under the auspices of the Bureau of American Ethnology. Attention focused upon the description and classification of native Amerindian

(i.e., non-Indo-European) languages – a source of material unavailable in Europe. The study of these languages was also in part motivated by missionary activities and Bible translation – enterprises that played no comparable role in Europe. As early as 1838, the French scholar P. E. Duponceau (1760–1844) argued that all Amerindian languages had a basic grammatical structure unknown in the Indo-European family, and popularized the term ‘polysynthetic’ to identify it. In 19th-century language typologies, polysynthetic languages were often distinguished from ‘synthetic’ and ‘analytic’ languages. The characteristic of polysynthetic languages was said to be a tendency to prefer long, complex words that ‘incorporate’ a variety of grammatical distinctions. Thus, for example, in Eskimo a single word expresses the idea that might be rendered in English by the sentence “Do you think he really intends to go to look after it?” (The term ‘incorporating’ is also used to describe this type of structure.)

One result of this American anthropological tradition of language study was the publication of the *Handbook of American Indian languages* (1911–1922). The team responsible for this was led by F. Boas (1858–1942), a German scholar who had emigrated to the United States. The *Handbook* comprised brief descriptive accounts of the phonetics, grammar, and vocabulary of a range of Amerindian languages. Boas himself rejected the view that all Amerindian languages were ‘polysynthetic.’ His most important theoretical contribution to general linguistics is sometimes said to be his observation that Europeans were prone to misdescribe the sounds of Amerindian languages, because they could not avoid hearing them through the grid imposed by the phonetics of the European languages they themselves spoke. This idea has obvious links to Saussure’s insistence that every language is structurally unique, although Boas arrived at this notion independently of Saussure.

In Europe, the chief influence of anthropology on the development of linguistics is to be seen in the work of the Polish anthropologist B. K. Malinowski (1884–1942), whose fieldwork in the Trobriand Islands inspired his view of language. Malinowski popularized the notion of “phatic communion,” which, in contrast to the traditional idea that language existed to communicate ‘thoughts,’ emphasized the role of language as a mode of action for establishing social bonds.

Malinowski’s ideas were taken up by J. R. Firth (1890–1960), professor of general linguistics at London University, whose ideas were in turn developed by ‘neo-Firthians,’ such as M. A. K. Halliday (b. 1925). Firth’s ‘polysystemic’ linguistics emphasized the need for linguists to set up different ‘systems’

in order to describe relations between linguistic units at different 'levels.' Firth viewed linguistics as language "turned back upon itself." He is sometimes regarded as the archetypal 'hocus-pocus' linguist, for whom linguistics is a verbal game played in accordance with metalinguistic rules devised by the linguist (as distinct from 'God's-truth' linguists, who view their task as being to discover some external linguistic reality that exists independently of their inquiries).

Linguistic Relativity

Another line of thinking in which Saussurean ideas link up with those emanating from the American anthropological tradition of language study concerns the theory that came to be known as 'linguistic relativity' or 'linguistic relativism.' It is also known as the 'Sapir-Whorf hypothesis,' of which there are various versions.

The term *relativity* is often taken as alluding to the work of A. Einstein (1879–1955) in physics. In one sense, this is not misguided, inasmuch as Einsteinian relativity emphasized the dependence of 'facts' on the viewpoint of the observer. The Sapir-Whorf hypothesis does likewise. Einstein himself, however, was far from being a linguistic relativist. Another figure often invoked in connection with the Sapir-Whorf hypothesis is the German polymath W. von Humboldt (1767–1835), whose ideas are seen as anticipating the hypothesis, in that he associated each national language with a distinctive way of viewing the world.

E. Sapir (1884–1939) was a student of Boas's and published a study of Takelma in the *Handbook of American Indian languages*. B. L. Whorf (1897–1941) worked for the Hartford Fire Insurance Company and became a student of Sapir's at Yale. In his insurance work, Whorf was struck by the fact that accidents were often caused by the way things and situations were described. (For example, an oil drum described as 'empty' had actually been full of highly dangerous vapor.) In its most general version, the Sapir-Whorf hypothesis claims that the language we speak provides a linguistic lens through which we view the world. It follows from this that speakers of different languages see the world – and interpret their own experience of it – in different, incommensurable ways.

This is probably the most revolutionary idea to emerge from modern linguistics – revolutionary in the sense that it overturns the assumptions underlying the development of language studies throughout the Western tradition. It also has far-reaching implications for linguistic epistemology. These implications tend to be ignored in mainstream linguistics, because they subvert the whole basis of establishing linguistics

as a 'science.' Insofar as mainstream linguistics has an answer to the problems raised by the Sapir-Whorf hypothesis, it lies in mounting a rival hypothesis claiming that there exists at some 'deep' level in human psychology a universal grammar common to all languages (Chomsky, 1986). It is difficult to see how this conflict of views could ever be resolved by the adduction of empirical linguistic 'evidence.' The debate is ultimately sterile. Both the Sapir-Whorf hypothesis and the universal grammar hypothesis open up routes that lead linguistics nowhere. That has not prevented partisans on both sides from devoting much time and effort to driving their own hypothesis further into no-man's-land.

Behaviorism

In America Saussure's work remained for some time without noticeable influence, in part because Saussurean linguistics eventually found itself under attack from psychologists, and in particular from linguists who had adopted the psychological doctrines of behaviorism. The leading linguistic exponent of these views between World War I and World War II was L. Bloomfield (1887–1949).

Although Bloomfield had himself studied Amerindian languages, and early in his career had been attracted to the psychological theories of Wundt, he was 'converted' to behaviorism at Ohio State University by A. P. Weiss (1879–1931). Bloomfield said of Weiss that he "was not a student of language, but he was probably the first man to see its significance" – an astonishing statement and an astonishing tribute. From the time of his acquaintance with Weiss, Bloomfield became increasingly and implacably opposed to 'mentalism' in linguistics. This entailed a rejection of Saussurean linguistics, given Saussure's view of the linguistic sign as a purely psychological unit existing in the human mind as a pairing of form (*signifiant*) with meaning (*signifié*). Somewhat more equivocal was Bloomfield's attitude to Boas's notion that linguistic structure lay below the level of human consciousness. Boas's introductory essay to the *Handbook of American Indian languages* had contained a concluding section on the "Unconscious Character of Linguistic Phenomena."

In accordance with behaviorist precepts, Bloomfield set out to banish from 'scientific' linguistics all appeal to unobservables. This meant no great loss as far as sound was concerned, since speech was audible and recordable. The problem lay with meaning. If meaning were in the mind of the speaker, they were unobservable and unrecordable. Accordingly, Bloomfield had to fall back on the ancient reocentrism of the Western tradition and declare that, for instance,

the meaning of the word *salt* is nothing other than the physical substance salt. Thus meaning was rescued and brought within the domain of linguistic observation. The question then arose, however, 'What is salt?' Recognizing that linguists could not be expected to be omniscient, Bloomfield's answer was that meanings could not be stated precisely except in domains where science had 'determined' the nature of the thing in question. Thus science had determined that salt in the physical world is sodium chloride. So the semantics of words for substances like salt posed no problem. But science had not yet determined what items in the 'real world,' if any, corresponded to such words as *love* and *hate*. There, according to Bloomfield, semantics would have to wait until scientists had got round to investigating such matters. So semantics was "the weak point in language-study" and was likely to remain so "until human knowledge advances very far beyond its present state" (Bloomfield, 1933).

The effect of this behaviorist doctrine was to skew linguistic inquiry in the direction of phonetics, since so much of semantics lay in the limbo of the unknown or unknowable.

Distributionalism

Some of Bloomfield's followers drew the conclusion that a 'science' of language was possible only if it ignored the analysis of meanings and concentrated solely on the analysis of linguistic forms. The attempt to theorize this nonsemantic approach to linguistic structure came to be regarded as characteristic of American (as opposed to Saussurean) 'structuralism.' Saussure would have regarded any such development as retrograde, rather than as an advance in the science of linguistics. In effect, it was a reversion to Greek atomism, although few if any of the American linguists involved were sufficiently well educated in the Classics to see the connection. Be that as it may, the outcome was a profound and misleading difference between the terms *structural* and *structuralism* as used in linguistics on either side of the Atlantic. Thus, for example, *Structural linguistics* (Harris, 1951) was the title of a well-known American textbook that (deliberately) excluded any consideration of linguistic meaning.

For the distributionalists, linguistic description consisted in identifying phonological units and defining all higher-order complexes by reference to the distribution of the basic units already defined. They assumed this could be done 'formally' – that is, without appealing at any stage to the supposed meaning of the items under investigation. The doyen of distributional linguistics was Z. S. Harris (1909–1992), whose critics argued that distributional analysis,

contrary to its proclaimed principles, tacitly took meaning into consideration at every turn. In other words – so the argument went – the linguist's task of determining whether two sounds were the same or different was impossible without information about whether the native speaker treated them as articulating a minimal semantic distinction.

Distributionalism was 'formalist' in another sense. It reduced the definition of linguistic units to the sum total of analytic procedures involved in identifying them. Adoption of a different set of procedures might produce a different set of units. This in turn had implications for the status of linguistics as a 'science.' (The assumption was that 'scientific' knowledge should not be at the mercy of the methodological preferences of the scientist.)

However, at a deeper level, distributionalism also connected with the revolution in the epistemology of science associated with Einsteinian relativity. In particular, the theory of operationalism in physics, as advanced by P. W. Bridgman (1882–1962), could be seen as a counterpart to Harris's distributionalism in linguistics. For Bridgman, all physical concepts were defined ultimately by reference to the techniques of measurement deployed in any scientific observation.

Generativism

One of Harris's students, A. N. Chomsky (b. 1929), proposed a novel approach to linguistic analysis by borrowing an idea from mathematical logic. It is possible to define a formal logistic system by setting out a string of 'rules' or algorithms that constrain transformations from one sequence of symbols to another. Chomsky saw that this procedure could be applied *in extenso* to the description of languages like English, if it were possible to propose a nucleus of symbols and transformational rules that would eventually 'generate' all and only the sentences of English. Many years of work were devoted to this Herculean quasi-mathematical labor, which never came to completion, either for English or for any other language.

Chomsky attacked behavioristic approaches to language, in particular the explanations offered by the experimental psychologist B. F. Skinner (1904–1990) concerning language acquisition. In the view of Chomsky and other generativists, language users were to be credited with reliable linguistic 'intuitions' as to the grammaticality or otherwise of expressions in their language. How these intuitions were to be distinguished from the products of teaching, rationalization, or personal preference was never clearly explained.

Although Chomsky at one point saw his project as a formalization of the Saussurean distinction between *langue* (= Chomskyan ‘competence’) and *parole* (= Chomskyan ‘performance’), that would have amused Saussure, for whom the naïve notion that a language could be reduced to a set of sentences was not even a starter. Generativism has been criticized by sociolinguists (see the next section) for ignoring linguistic variation and verbal interactions in ‘real-world’ contexts. In particular, Chomsky’s frequent appeal to an “ideal speaker-listener” is seen as an unacceptable evasion of linguistic issues that should be addressed empirically.

Sociolinguistics and Pragmatics

During the 20th century, there had also emerged in linguistics the simple but powerful idea that languages could be studied independently of any theory about “the language system.” So it did not matter too much what theorists like Saussure or Bloomfield or Chomsky said. Whatever they said, linguistic behavior was still available for description as a social phenomenon. This is the basic idea behind what came to be called ‘sociolinguistics.’

It derives from traditional dialectology, but it now incorporates the notion that linguistic variation has other than geographical parameters. In particular, it embraces the notion that there are discernible patterns of linguistic variation that depend on age, sex, education, social status, social occasion, professional occupation, and other factors. All these may intersect in numerous ways. It is the task of the sociolinguist to document, disentangle, and explicate these complex patterns.

A sociolinguistic approach leads on naturally to ‘pragmatics’ or ‘pragmalinguistics.’ This has now become a catch-all category embracing virtually all forms of linguistic variation or distinguishing characteristics that can be detected when language is used in a particular type of activity or communication situation.

The main objection that has been leveled against linguistic research of this kind is that it falls inevitably under the aegis or influence of sociology, and thus is led to adopt techniques of investigation and methods of classification that are approved in that discipline (in particular the adoption of certain social classifications and statistical assessments that may perhaps be inappropriate or misleading for linguistic purposes).

Speech-Act Theory

Pursuing a sociolinguistic approach to its logical conclusion, one is led to ask not “What are the

basic sound units of a language?”, nor “What are the basic grammatical units and constructions?”, but “What are the basic speech acts by which one human being communicates with another?” Curiously, this question had been more or less neglected in modern linguistics until it was raised in the neighboring discipline of philosophy. That it was raised at all is due largely to the Oxford philosopher J. L. Austin (1911–1960), whose work in turn owes much to the climate of inquiry created in philosophy of language and philosophy of mathematics by G. Frege (1848–1925), B. A. W. Russell (1872–1970), G. E. Moore (1873–1958), and L. Wittgenstein (1889–1951).

The title of Austin’s brief but highly influential book *How to do things with words* (1962) strongly suggests the pragmatic orientation of his thought. He realized that stating a fact (recognized in traditional grammar as the province of the indicative mood) was only one of many verbal activities it was possible to engage in, even though it was the activity that traditional philosophy valued and paid attention to above all others. This led Austin to ask what other types of act were available to the speaker or writer. He pointed out that saying “I apologize” is not a description of an apology but the execution or performance of the act in question. This simple observation had escaped all traditional grammarians, for whom *I apologize* is on a par with any other first-person present indicative active, such as *I walk*.

Austin’s initial exposition of the performative dimension in linguistic communication was followed up in work by J. Searle (b. 1932) and H. P. Grice (1913–1988). No comprehensive taxonomy of speech acts has been forthcoming. One reason alleged by critics of speech-act theory is that this is impossible, because speech-acts proliferate *ad infinitum*, depending on the criteria adopted for identifying them. Another, perhaps more damaging, criticism is that speech-act theory itself tacitly relies on accepting of the prior existence of a code (i.e., a language) by means of which the act is articulated. In other words, the speech-act theorist is led into complicity with the traditional assumptions of orthodox linguistics. Without these, the speech act cannot be identified or described.

Language Origins and Nonhuman Language

The ancient question of linguistic origins fell out of favor in the course of the 19th century. The Société de Linguistique, founded in Paris in 1866, expressly forbade the presentation of papers on the subject. The eminent American philologist W. D. Whitney

(1827–1894) dismissed all discussion of the origins of language as “mere windy talk.” Interest was renewed, however, in the 20th century, largely through research from two directions.

With the general acceptance of Darwin’s theory of evolution, it seemed likely to many that language itself had developed from more primitive forms of communication used by less advanced ancestors of the human race. On this view, it ought to be possible in principle to account for language as having evolved on the basis of natural selection.

Others thought it more probable that language was the result of a unique genetic mutation. Others again maintained the skepticism of the 19th century and insisted that, in the present state of knowledge, it is impossible to progress beyond speculation to ‘science.’ The debate continues (Botha, 2003).

In the study of animal behavior, much attention was paid in the second half of the 20th century to the possibility of teaching linguistic skills to other primates, particularly those genetically close to *Homo sapiens*. It was generally accepted that apes were, for physiological reasons, unlikely to be able to produce articulated speech. But there seemed no reason why they might not master a language of gestures. Some of these projects developed systems based on American Sign Language for the Deaf. In later experiments, computer-monitored keyboards were used, and apes proved adept at mastering them.

From the beginning, work in this area was controversial. Skeptics questioned whether the use made by apes of the signs taught to them by human instructors was genuinely ‘linguistic,’ as distinct from being merely ‘imitative.’ (Pigeons had been taught to ‘play ping-pong,’ or at least to go through the motions, but no one seriously supposed that the birds understood what they were doing – i.e., playing a game.) However, it subsequently emerged that bonobos not only could use signs regularly to communicate their own wants, but could understand spoken English of a certain degree of grammatical complexity (Savage-Rumbaugh *et al.*, 1998).

Retrospect

When one looks back at the ambitious academic program that Saussure (1916) first laid out for modern linguistics, it is interesting to note what progress has been made. The program had three objectives. There is still a long way to go in pursuing Saussure’s first goal of describing all known languages and recording their history. His second goal, of determining the forces operating permanently and universally in all languages, is no nearer to fulfillment, because of fundamental disagreement about the methods to be

pursued. His third goal, that of defining and delimiting linguistics itself, threatens to disappear as the boundaries separating linguistics from neighboring disciplines become increasingly porous. The real weakness in Saussure’s original program, as it now appears, was its failure to make provision for any study of the actual contexts in which linguistic communication is embedded. Such studies seem inevitably to involve trespassing on or borrowing from work done by specialists in adjacent fields. As a result, fewer and fewer linguists any longer believe in the autonomy of linguistics. It would be a more accurate reflection of current trends to say that across a wide spectrum of academic research in the humanities the importance of linguistic considerations is increasingly acknowledged, and the interdependence between verbal and nonverbal communication recognized.

See also: Behaviorism: Varieties; Speech Acts.

Bibliography

- Andresen J T (1990). *Linguistics in America 1769–1924: a critical history*. London: Routledge.
- Ardener E (ed.) (1971). *Social anthropology and language*. London: Tavistock.
- Asher R E (ed.) (1994). *The encyclopedia of language and linguistics*. Oxford: Pergamon.
- Austin J L (1962). *How to do things with words*. Oxford: Clarendon.
- Bloomfield L (1933). *Language*. New York: Holt.
- Boas F (ed.) (1911–1922). *Handbook of American Indian languages*. Washington, D. C.: Government Printing Office.
- Botha R (2003). *Unravelling the evolution of language*. Oxford: Elsevier.
- Bréal M (1897). *Essai de sémantique*. Paris: Hachette.
- Chomsky A N (1986). *Knowledge of language*. New York: Praeger.
- Davis P W (1973). *Modern theories of language*. Englewood Cliffs: Prentice-Hall.
- Fries C C (1966). *Linguistics: the study of language*. New York: Holt, Rinehart & Winston.
- Harris R (ed.) (1988). *Linguistic thought in England 1914–1945*. London: Duckworth.
- Harris R (ed.) (1995). *Language and linguistics* (4 vols). London: Routledge/Thoemmes.
- Harris R (ed.) (1997). *Origins of American linguistics 1643–1914* (13 vols). London: Routledge/Thoemmes.
- Harris R (2003). *Saussure and his interpreters* (2nd edn.). Edinburgh: Edinburgh University Press.
- Harris Z S (1951). *Structural linguistics*. Chicago: University of Chicago Press.
- Jespersen O (1922). *Language: its nature, development and origin*. London: Allen & Unwin.
- Joseph J E, Love N & Taylor T J (2001). *Landmarks in linguistic thought II: the Western tradition in the twentieth century*. London: Routledge.

- Lepschy G C (1970). *A survey of structural linguistics*. London: Faber & Faber.
- Mohrmann C, Sommerfelt A & Whatmough J (eds.) (1961). *Trends in European and American linguistics 1930–1960*. Utrecht: Spectrum.
- Robins R H (1994). 'Jones, Sir William (1746–94).' In Asher R E (ed.) *The encyclopedia of language and linguistics*. Oxford: Pergamon. 1820–1821.
- Saussure F de (1916). *Cours de linguistique générale*. Paris: Payot.
- Savage-Rumbaugh S, Shanker S G & Taylor T J (1998). *Apes, language and the human mind*. New York: Oxford University Press.
- Vachek J (1966). *The linguistic school of Prague: an introduction to its theory and practice*. Bloomington: Indiana University Press.

Modularity

R Carston, University College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

Here is a general characterization of the concept of a mental module: it is a department of the mind that is dedicated to a specific domain of content and is largely autonomous from other departments of the mind. Possible content domains are visual perception, language, arithmetic, music, spatial orientation, social relations, and scientific reasoning, each of which may itself be made up of more fine-grained domains.

Although domain specificity and autonomy are central to any notion of modularity, there are also important differences in the kinds of mental structure that may be referred to as modules. The main distinction is between (a) systems of mental representations and (b) computational mechanisms (see Samuels, 2000). Modules of the first sort are sometimes called 'information' modules, 'Chomskyan' modules (after Noam Chomsky, who proposed such a modular view of the human language capacity), or 'competence' modules. Modules of the second sort are computational devices, processing systems, or 'performance' modules. Of course, modular systems of both kinds often coexist. For instance, a modular system for performing arithmetic computations might call on a domain-specific system of arithmetic knowledge; similarly, a modular mechanism for comprehending language (a parser) might deploy the contents of a Chomskyan module of linguistic information (see Segal, 1996).

Chomskyan modules ('C-modules' hereafter) are domain-specific bodies of knowledge that underpin a cognitive capacity. In addition to the well-known case of linguistic knowledge (the phonological, syntactic, and semantic knowledge constituting an individual's language), plausible candidates for such competence modules are physical knowledge (knowledge of some of the properties of physical objects),

biological knowledge (knowledge of some of the properties of living things), psychological knowledge (knowledge of some of the properties of human minds), and mathematical knowledge (knowledge of counting and numerical quantities).

However, there are bodies of domain-specific knowledge that do not qualify as C-modules. Consider, for instance, a child's beliefs about Batman or an adult's knowledge of British history. To count as a C-module, a knowledge system must meet two additional constraints. First, there must be restrictions on information flow (both the accessibility of the system's knowledge to other content domains and the effect of outside information on the system itself). Second, at least some subset of the knowledge comprising a C-module is innately given (hence universal). With regard to language, a strong case has been made for a universal grammar (UG), an innate system of mental representations, which is shaped by environmental input into a specific linguistic system (Chomsky, 1984; Chomsky, 1986). By analogy with UG, we can envision that each of the other possible C-modules mentioned above develops from an innate knowledge base: UPh, UBio, and so on.

Computational or performance modules ('P-modules' hereafter) are mechanisms that process and transform mental representations. They are only able to carry out computations on a restricted range of inputs, namely, those concerning a particular domain, and in processing these inputs they can only use their own proprietary principles and procedures (their own database, which may comprise a C-module). For instance, an arithmetic computational module will only provide solutions to arithmetic problems and will only use arithmetic operations to solve those problems. Fodor's landmark 1983 book on modularity focused on P-modules (while acknowledging the existence of C-modules), arguing in particular that language processing is a P-modular system.

In addition to the central properties of domain specificity and informational encapsulation, Fodor

suggested a number of other criterial properties of P-modules. Like reflexes, they function mandatorily when presented with an appropriate input, and they are very fast in their operation; consequently, their output is relatively shallow. There are three criteria that reflect their innate basis, distinguishing them from learned habits (which may also be fast and automatic): neural localization, a relatively fixed ontogenetic pace and sequence, and characteristic patterns of independent breakdown. Fodor (1983) provides evidence that language parsing (or perception) has all eight of these properties, though others dispute this (Marslen-Wilson and Tyler, 1987).

The big question about the architecture of the human mind concerns the extent to which it is modular as opposed to employing domain-general principles and processes. Fodor's (1983, 1989, 2000) answer is that only input and output systems (roughly, the various perceptual systems, motor systems, and language) are modular. The central cognitive systems, responsible for belief fixing and reflective thought, are content general and highly sensitive to contexts and individual utilities. So, according to Fodor, they are nonmodular (global, isotropic) and, as a result, are beyond the reach of scientific investigation.

The strict Fodorian view of a module, with its application only to peripheral mental systems, is widely seen as too restrictive (Karmiloff-Smith, 1992; Sperber, 1996, 2005; Carruthers, 2003, 2004). From an evolutionary perspective on human cognition (Cosmides and Tooby, 1994; Sperber, 1996), what characterizes a mental module is that it is a special-purpose mechanism attuned to regularities in a particular domain, a biological adaptation selected for its efficiency in dealing with problems specific to that domain. It will tend to employ principles and heuristics that have evolved for the task at hand and are not applicable to problems outside its domain.

On this broader view of a mental module, human cognitive architecture is likely to be massively modular. Not only perception but also many central conceptual systems are modular (Sperber, 1996, 2002; Carruthers, 2003). One central inferential capacity that has received much attention as a likely modular system is 'mind reading,' that is, the capacity to attribute mental states such as intentions and beliefs to others on the basis of observed behavior (Scholl and Leslie, 1999; Wilson, 2003). It is further claimed that the pragmatic capacity (that is, the ability to infer the intended meaning of communicative acts, whether verbal or nonverbal) is a sub-module of the mind-reading module, with its own special-purpose principles and mechanisms (Sperber and Wilson, 2002; Wilson, 2003).

Strong supporting evidence for any modularity claim comes from 'double dissociations,' where for one group of people a specific capacity (say, mind reading) is impaired while another (say, language) remains intact, and, for another group, the pattern is the reverse (see Shallice, 1988; Smith and Tsimpli, 1996; Wilson, 2003).

Bibliography

- Cain M (2002). 'The modularity thesis.' In Cain M (ed.) *Fodor*. Polity Press. Chap. 7.
- Carruthers P (2003). 'Moderately massive modularity.' In O'Hear A (ed.) *Minds and persons*. Cambridge: Cambridge University Press.
- Carruthers P (2004). 'The mind is a system of modules shaped by natural selection.' In Hitchcock C (ed.) *Contemporary debates in the philosophy of science*. Oxford: Blackwell.
- Carston R (1996). 'The architecture of the mind: modularity and modularization.' In Green D *et al.* (eds.) *Cognitive science: an introduction*. Oxford: Blackwell. 53–83.
- Chomsky N (1984). *Modular approaches to the study of the mind*. San Diego: San Diego State University Press.
- Chomsky N (1986). *Knowledge of language: its nature, origins, and use*. New York: Praeger.
- Cosmides L & Tooby J (1994). 'Origins of domain-specificity: The evolution of functional organization.' In Hirschfeld L & Gelman S (eds.) *Mapping the mind: domain specificity in cognition and culture*. Cambridge: Cambridge University Press. 85–116.
- Fodor J (1983). *Modularity of mind*. Cambridge: MIT Press.
- Fodor J (1989). 'Why should the mind be modular?' In George A (ed.) *Reflections on Chomsky*. Oxford: Blackwell. 1–22.
- Fodor J (2000). *The mind doesn't work that way*. Cambridge: MIT Press.
- Gallistel R (2000). 'The replacement of general-purpose learning models with adaptively specialized learning modules.' In Gazzaniga M (ed.) *The new cognitive neurosciences*, 2nd edn. Cambridge: MIT Press.
- Karmiloff-Smith A (1992). *Beyond modularity: a developmental perspective on cognitive science*. Cambridge: MIT Press.
- Marslen-Wilson W & Tyler L (1987). 'Against modularity.' In Garfield J (ed.) *Modularity in knowledge representation and natural-language understanding*. Cambridge: MIT Press. 37–62.
- Pinker S (1994). *The language instinct*. London: Allen Lane, The Penguin Press.
- Pinker S (1997). *How the mind works*. New York: Norton.
- Samuels R (2000). 'Massively modular minds: evolutionary psychology and cognitive architecture.' In Carruthers P & Chamberlain A (eds.) *Evolution and the human mind: modularity, language and meta-cognition*. Cambridge: Cambridge University Press. 13–46.
- Scholl B & Leslie A (1999). 'Modularity, development and "theory of mind."' *Mind and Language* 14, 131–153.
- Segal G (1996). 'The modularity of theory of mind.' In Carruthers P & Smith P (eds.) *Theories of*

- theories of mind*. Cambridge: Cambridge University Press. 141–155.
- Shallice T (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.
- Smith N & Tsimpli I-M (1996). 'Modules and quasi-modules: language and theory of mind in a polygot savant.' *UCL Working Papers in Linguistics* 8, 1–17.
- Sperber D (1996). 'Mental modularity and cultural diversity.' In Sperber D (ed.) *Explaining culture: a naturalistic approach*. Oxford: Blackwell. 119–150.
- Sperber D (2002). 'In defense of massive modularity.' In Dupoux E (ed.) *Language, brain and cognitive development: essays in honor of Jacques Mehler*. Cambridge: MIT Press. 47–57.
- Sperber D (2005). 'Massive modularity and the first principle of relevance.' In Carruthers P, Laurence S & Stich S (eds.) *The innate mind: structure and content*. Cambridge: Cambridge University Press.
- Sperber D & Wilson D (2002). 'Pragmatics, modularity and mind-reading.' *Mind & Language* 17, 3–23.
- Wilson D (2003). 'New directions for research on pragmatics and modularity.' *UCL Working Papers in Linguistics* 15, 105–127.

Monotonicity and Generalized Quantifiers

R Cann, University of Edinburgh, Edinburgh, UK

© 2006 Elsevier Ltd. All rights reserved.

Many quantified noun phrases give rise to entailment relations between sentences that differ with respect to some modification of a verb phrase in their scopes. Consider the following sentences:

- (1a) No students laughed.
- (1b) No students laughed loudly.
- (2a) Every philosopher smokes.
- (2b) Every philosopher smokes heavily.

Whereas Example (1a) entails Example (1b), this is not the case with Examples (2a) and (2b). On the other hand, Example (2b) *does* entail Example (2a), whereas Example (1b) does not entail Example (1a). The important property to note here is that in the (b) sentences, the verb phrases pick out a smaller set of entities than the (a) sentences do (those who laugh loudly constitute a subset of those who laugh, and the set of heavy smokers is a subset of the set of smokers). So, the truth value of a sentence with *No students* in its subject position is unaffected if the denotation of its verb phrase (VP) is reduced. At the same time, the truth value of a sentence with *every philosopher* as subject remains unaffected whenever the denotations of its VP is extended. These entailment relations are known as 'monotonicity' relations, and *no students* is said to be 'monotone decreasing' (MON↓); *every philosopher* is 'monotone increasing' (MON↑). So a quantified noun phrase (QNP) is monotone just in case:

$$\text{QNP VP}_1 \vdash \text{QNP VP}_2,$$

where \vdash is the entailment relation and the denotation of VP_1 contains that of VP_2 (decreasing) or vice versa

(increasing). Some further examples of monotone-increasing quantifiers are shown in the following examples:

- (3a) Proper names:
Kim sang loudly \vdash Kim sang.
- (3b) *Many N*:
Many lecturers eat a lot of chocolate \vdash
Many lecturers eat chocolate.
- (3c) *At least n N*:
At least three cats were killed by the dog \vdash
At least three cats were killed.
- (3d) *Most N*:
Most students passed with high marks \vdash
Most students passed.
- (3e) *More than half of the N*:
More than half of the congregation came to church on time \vdash
More than half of the congregation came to church.

There are fewer monotone-decreasing quantifiers that do not directly involve a negation particle:

- (4a) *Few N*:
Few lecturers eat chocolate \vdash
Few lecturers eat a lot of chocolate.
- (4b) *At most n N*:
At most three cats were killed \vdash
At most three cats were killed by the dog.
- (4c) *Less than half of the N*:
Less than half of the congregation came to church \vdash
Less than half of the congregation came to church on time.

Not all quantifiers are monotonic, however (represented as MON~):

- (5a) *Exactly n N*:
Exactly three cats were killed \nvdash
Exactly three cats were killed by the dog.

(5b) *A few N*:

A few lecturers eat chocolate ~~not~~

A few lecturers eat a lot of chocolate.

Interestingly enough, it seems to be universally the case that ‘simple’ noun phrases – i.e., proper names, noun phrases consisting of a simple determiner plus noun, and quantifier noun phrases such as *everyone*, *no one* – are always monotonic. Only noun phrases containing complex determiners, such as *a few*, *exactly five*, *all but seven*, etc., may not be.

To give more precise definitions and to explore the properties of monotonicity, most semanticists follow Montague (1973) in taking QNPs as denoting **generalized quantifiers**. These are defined as higher order denotations, picking out sets of sets of entities, each set bearing some specified relation to the set of entities denoted by the common noun (phrase) in the QNP. Thus, *every philosopher* denotes the set of all sets that contain as a subset the set of philosophers. *Every philosopher smokes* expresses a true proposition, therefore, just in case the set of smokers is a superset of the set of philosophers, in which case $\| \text{smoke}' \| \in \| \text{every philosopher} \|$ (where $\| \alpha \|$ represents the denotation of α). More generally, given a sentence consisting of a quantified noun phrase in subject position and some verb phrase, $\| \text{QNP VP} \|$ is then defined as being true just in case $\| \text{VP} \| \in \| \text{QNP} \|$.

The monotonicity properties of generalized quantifiers have been the object of considerable research within semantics since the seminal work of Barwise and Cooper (1981). Within this framework, the monotonicity property of a quantifier is easily defined. A generalized quantifier QNP VP is as follows:

(6a) MONOTONE INCREASING just in case whenever $\| \text{VP}_1 \| \in \| \text{QNP} \|$ then $\| \text{VP}_2 \| \in \| \text{QNP} \|$ where $\| \text{VP}'_1 \| \subseteq \| \text{VP}'_2 \|$.

(6b) MONOTONE DECREASING just in case whenever $\| \text{VP}'_1 \| \in \| \text{QNP} \|$ then $\| \text{VP}_2 \| \in \| \text{QNP} \|$ where $\| \text{VP}_2 \| \subseteq \| \text{VP}'_1 \|$.

Quantified noun phrases may combine with the ‘Boolean’ connectives *not*, *and*, *or*, and these combinations may have an effect on the monotonicity of the resulting quantifier. For example, negation reverses the direction of the monotonicity of a quantifier, whether it is internal (corresponding to VP negation) or external (corresponding to the negation of the QNP).

(7a) Not all philosophers smoke \vdash Not all philosophers smoke heavily.

(7b) All students don’t like semantics \vdash All students don’t like formal semantics.

(7c) Few lecturers don’t eat a lot of chocolate \vdash Few lecturers don’t eat chocolate.

Formally, negation of quantifiers can be defined as follows:

External negation :

$$\| \neg \text{QNP} \| = \{ X \subseteq E \mid X \notin \| \text{QNP} \| \}. \quad (8a)$$

Internal negation :

$$\| \text{QNP} \neg \| = \{ X \subseteq E \mid (E - X) \in \| \text{QNP} \| \}. \quad (8b)$$

To see the effects of these definitions, consider the negation of the quantifier *all N*:

$$(9a) \| \text{all N} \| = \{ X \subseteq E \mid \| \text{N} \| \subseteq X \} = \{ X \subseteq E \mid \| \text{N} \| \cap X = \| \text{N} \| \}.$$

$$(9b) \| \neg \text{all N} \| = \{ X \subseteq E \mid X \notin \| \text{all N} \| \} = \{ X \subseteq E \mid \| \text{N} \| \cap X \neq \| \text{N} \| \} \\ \text{(external negation plus (9a))}.$$

$$(9c) \| \text{all N} \neg \| = \{ X \subseteq E \mid (E - X) \in \| \text{all N} \| \} \\ = \{ X \subseteq E \mid \| \text{N} \| \cap (E - X) = \| \text{N} \| \} \\ = \{ X \subseteq E \mid \| \text{N} \| \cap X = \emptyset \} \\ = \| \text{no N} \| \\ \text{(internal negation plus (9a))}.$$

To see that negation does indeed reverse monotonicity, consider how the definitions in Examples (8a) and (8b) affect a monotone increasing quantifier. Assume a QNP is $\text{MON}\uparrow$ and that X and Y are arbitrary sets such that $X \subseteq Y$, so that ‘QNP X ’ entails ‘QNP Y ’ (where X and Y are VPs that denote X and Y , respectively). If $Y \in \| \text{QNP} \neg \|$, then $(E - Y) \in \| \text{QNP} \|$ (by Example (8b)) and $(E - Y) \subseteq (E - X)$, so $(E - X) \in \| \text{QNP} \|$ and hence $X \in \| \text{QNP} \neg \|$. So ‘QNP not Y ’ entails ‘QNP not X ,’ as required. Analogously, if $Y \in \| \neg \text{QNP} \|$, then $Y \notin \| \text{QNP} \|$, so that $X \notin \| \text{QNP} \|$ and hence $X \in \| \neg \text{QNP} \|$. So ‘not QNP Y ’ entails ‘not QNP X ,’ and both forms of negation thus make a monotone-increasing quantifier into a decreasing one.

Note that because $\text{QNP} = \neg \neg \text{QNP} = \text{QNP} \neg \neg$, any decreasing simple quantifier is a negated form of a simple increasing one. This may explain why negating decreasing QNPs is generally ill formed, because the corresponding simple increasing QNPs could be used.

(10) *Not few students = Many students.

Monotonicity is maintained if two (or more) QNPs with the same direction of monotonicity are conjoined by *and* or *or*.

(11a) A student and every lecturer were snoring loudly \vdash

A student and every lecturer were asleep.

(11b) Bill or both women ran home quickly \vdash Bill or both women ran home.

(12a) No student and not all lecturers were asleep \vdash No student and not all lecturers were snoring loudly.

- (12b) At most ten men or less than half the women
ran home \vdash
At most ten men or less than half the women
ran home quickly.

Again it is fairly easy to show that this must be the case using the formal apparatus of generalized quantifiers, given conjunction interpreted as involving set intersection and disjunction as set union, as in Examples (13a) and (13b):

- (13a) $\parallel NP_1 \parallel$ and $NP_2 \parallel = \parallel NP_1 \parallel \cap \parallel NP_2 \parallel$.
(13b) $\parallel NP_1 \parallel$ or $NP_2 \parallel = \parallel NP_1 \parallel \cup \parallel NP_2 \parallel$.

We take the persistence of monotonicity properties with *and* as an illustration of how the proofs of these four theorems are established:

Assume that QNP_1 and QNP_2 are both $MON\uparrow$ and that X and Y are arbitrary sets such that $X \subseteq Y$. If $X \in \parallel QNP_1 \parallel \cap \parallel QNP_2 \parallel$, then it must be the case that $X \in \parallel QNP_1 \parallel$ and $X \in \parallel QNP_2 \parallel$. So, by increasing monotonicity, $Y \in \parallel QNP_1 \parallel$ and $Y \in \parallel QNP_2 \parallel$. Hence, $Y \in \parallel QNP_1 \parallel \cap \parallel QNP_2 \parallel$. QED.

Barwise and Cooper (1981) suggest that coordination of noun phrases is constrained by monotonicity:

Co-ordination constraint: Two NPs can be coordinated by conjunction (*and*) and disjunction (*or*) iff they are both monotone increasing or both monotone decreasing.

- (14a) No student and every lecturer liked the book.
(14b) Most students and every lecturer liked the book.
(14c) *No student and most lecturers liked the book.
(14d) *No student and a lecturer liked the book.

On the other hand, the connective *but* seems to require quantifiers with different monotonicity:

- (15a) Every woman but no man agreed with the Prime Minister.
(15b) No student but most lecturers liked the book.
(15c) *Most students but every lecturer liked the book.

However, there are many counterexamples to these apparent constraints, so they cannot be adopted universally and some other explanation remains to be given for the ill-formed examples.

- (16a) At least three cats and at most four dogs (were killed on this road last year).
 $MON\uparrow$ and $MON\downarrow$
(16b) Bill and a few women (met in the pub).
 $MON\uparrow$ and $MON\sim$:

Although the constraint against conjoining quantified noun phrases with different monotonicity properties cannot be maintained as a linguistic universal, there is one property that is fully robust. Negative polarity items such as *any* or *ever* are typically within the scope of negation (Examples (17a)–(17d) (Klima,

1964), but they also appear in the scope of certain types of quantified noun phrases that do not contain an overt negation marker (Examples (18a)–(18d)). The latter are all monotone-decreasing quantifiers and the generalization in Example (19) appears to be universally true (Ladusaw, 1983):

- (17a) John didn't eat any chocolate.
(17b) *John ate any chocolate.
(17c) No students were ever early.
(17d) *All students are ever early.
(18a) At most five residents saw anything.
(18b) Few students fail any exams.
(18c) *Many students fail any exams.
(18d) Less than half of the congregation ever stays to the end of the service.
(18e) *Exactly five people ever sing loudly enough.

(19) Negative polarity items may appear in the scope of monotone-decreasing quantifiers but not in the scope of nonmonotonic or monotone-increasing quantifiers.

In the discussion so far, the monotonicity of quantifiers have been treated as referring to entailment relations involving increases or decreases in their **scope**, as denoted by a VP with which they may combine. Because quantified noun phrases consist of a determiner plus a common noun phrase, the latter, like verb phrases, are taken in extensional semantic theories to denote sets; the theory of generalized quantifiers conceptualizes **determiners** as denoting relations between sets:

- (20) $DET(A) (B)$, where A is the set denoted by the common noun phrase (the restrictor) and B is the set denoted by the verb phrase (the scope).

Given this view, it is natural to consider whether monotonicity applies not only to variations in the scope of a determiner but also in its restrictor. Clearly, this is the case, and restrictor (or left) monotonicity can be defined in the same way as scope (or right) monotonicity was defined for QNPs (some examples are given in Example (24)).

- (21) A determiner DET is LEFT (OF RESTRICTOR) MONOTONE INCREASING ($\uparrow MON$) iff
 $DET N_1 VP \vdash DET N_2 VP$, where
 $\parallel N_1 \parallel \subseteq \parallel N_2 \parallel$.
A student with red hair is singing entails
A student is singing.
(22) A quantifier DET is LEFT (OF RESTRICTOR) MONOTONE DECREASING ($\downarrow MON$) iff
 $DET N_1 VP \vdash DET N_2 VP$, where
 $\parallel N_2 \parallel \subseteq \parallel N_1 \parallel$.
No student failed their degree exams entails
No student of linguistics failed their degree exams.
(23) A quantifier DET is LEFT (OF RESTRICTOR) NONMONOTONIC ($\sim MON$) iff

DET N_1 VP $\not\models$ DET N_2 VP, where
 $\|N_2\| \subseteq \|N_1\|$ or $\|N_1\| \subseteq \|N_2\|$.

Exactly five students failed badly neither entails
 nor is entailed by *Exactly five students of
 linguistics failed badly*.

- (24) \uparrow MON \uparrow A, some, at least n
 \downarrow MON \downarrow No, less than n
 \downarrow MON \uparrow Every, all
 \downarrow MON \uparrow Not all, not every
 \sim MON \uparrow Most, many
 \uparrow MON \sim Some but not all
 \sim MON \downarrow Few
 \sim MON \sim Exactly n

See also: Compositionality: Semantic Aspects; Formal Semantics; Quantifiers: Semantics.

Bibliography

Bach E, Jelinek E & Partee B (eds.) (1995). *Quantification in natural languages*. Dordrecht: Kluwer.

- Barwise J & Cooper R (1981). 'Generalized quantifiers and natural language.' *Linguistics and Philosophy* 4, 159–219.
 Gamut L T F (1991). *Logic language and meaning*. Chicago: Chicago University Press.
 Gärdenfors P (1987). *Generalized quantifiers: linguistic and logical approaches*. Dordrecht: Reidel.
 Keenan E L (1996). 'The semantics of determiners.' In Lappin S (ed.) *The handbook of contemporary semantics*. Oxford: Basil Blackwell.
 Keenan E L & Stavi J (1986). 'A semantic characterization of natural language determiners.' *Linguistics and Philosophy* 9, 253–326.
 Klima E (1964). 'Negation in English.' In Fodor J A & Katz J J (eds.) *The structure of language, readings in the philosophy of language*. Englewood Cliffs, NJ: Prentice-Hall. 246–323.
 Ladusaw W (1983). 'Logical form conditions on grammaticality.' *Linguistics and Philosophy* 6, 389–422.
 Montague R (1973). 'The proper treatment of quantification in English.' In Thomason R (ed.) *Formal philosophy*. New Haven: Yale University Press.
 van Benthem J (1986). *Essays in logical semantics*. Dordrecht: D. Reidel.

Montague Semantics

T M V Janssen, ILLC, University of Amsterdam, Amsterdam, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Historical Background

Montague semantics emerged around 1970, and constituted a fundamentally new approach to the semantics of natural language. In order to understand the importance of this approach, it is useful to consider the situation in that period of some neighboring disciplines of semantics: philosophy of language, philosophical logic, and linguistics. These subjects are considered below, a more extended picture of the historical background is given by Partee and Hendriks (1997).

One of the subjects studied in philosophy of language was meanings of natural language, and these were (before 1970) sometimes represented in some or other logic. The mapping between a sentence and its logical representation was made in an *ad hoc* way; it was more or less stipulated which formula was the correct meaning representation of a given sentence. The situation could be characterized as follows: a 'bilingual logician,' who knew logic and who knew a natural language, provided the formula. It was noticed that there could be a large difference between

the sentence and the formula, and a dominant view of these matters at that time was the so-called 'misleading form thesis,' saying that there is a sharp distinction between the grammatical and the logical form of sentences, and that the grammatical form disguises the logical form to such an extent that they cannot be related in a systematic way. It was sometimes even proposed to design for certain purposes an improved version of natural language in which the form does not obscure the meaning.

In philosophical logic there is a tradition to study words of natural language that are philosophically interesting. Examples are 'necessarily' and 'possibly.' Axiom systems for these notions were designed which expressed their properties. A jungle of systems of modal logics arose, motivated by different properties of embeddings of these notions. Kripke brought about, in the mid-1960s, an enormous change in this field. He introduced semantic models for modal logics, thereby making it possible to conceive modal logic in the same way as mathematical logic: as a formal language with a model-theoretic semantics. The variety of systems could be structured by conceiving them as expressing relations between possible worlds in a model.

Around 1960, Chomsky brought about great changes in linguistics by introducing mathematical standards of explicitness. He developed the tools for this (context-free grammars and transformations),

and syntax became a flourishing branch of linguistics. There was some attention to semantic issues, such as whether transformations were meaning-preserving, or what would be the input for a semantic component, but the theory was a theory about syntax that did not deal explicitly with semantics.

These three lines were brought together by Richard Montague. He was a mathematical logician who had made important contributions to the foundations of set theory. He was attracted by Chomsky's formal treatment of natural language, but unsatisfied by its (lack of) treatment of semantics. Therefore he developed an alternative to the Chomskyan approach that satisfied his (logical) standards. He presented a fragment of English and provided it with a rigorous model-theoretic interpretation. Most important is the fact that the relation between a sentence and its meaning is defined in a systematic way. It became possible, for the first time in history, to calculate which meaning is associated with any given sentence, hence to make predictions concerning semantics.

By developing his grammar model, Montague provided evidence for his opinion that there is no important theoretical difference between natural languages and the languages studied by logicians: both can be dealt with using the same methods and with the same mathematical rigor (Montague, 1970a: 189; Montague, 1970b: 313; Thomason, 1974: 188, 222). The title of one of Montague's first publications on this subject provides clear evidence of his position: 'English as a formal language' (Montague, 1970a).

Aims

The aim of Montague semantics is to describe, predict, and explain semantic relations between natural language utterances, an aim it shares with other theories. In the present section, some important semantic relations between natural language utterances will be introduced by means of examples, viz. entailment, valid reasoning, synonymy, and ambiguity. The examples given here are realistic in the sense that they are treated within the field of Montague semantics, thus giving an impression of the variety of phenomena that are studied. The examples that occur without reference are within the fragment of Montague (1973), or are variants of such examples.

An important semantic relation is the 'entailment' relation between sentences, say A and B. This means that whenever A is accepted as being true, B must also be accepted as being true, on grounds of meaning properties. Sentence (1) entails sentence (2):

- (1) Mary is singing and dancing.
- (2) Mary is singing.

This entailment, however, does not hold for all grammatical subjects: witness (3) and (4), where in fact the inverse relationship holds.

- (3) No-one is singing and dancing.
- (4) No-one is singing.

This means that the noun phrases have to be divided into two classes, one for which the entailment holds, and one for which it does not. Then one would like to have an explanation of why precisely *two girls* and *both girls* are in different classes (*Both/precisely two girls were singing and dancing*), and a prediction concerning compound terms like *few girls and many boys*. For an overview of properties of quantified noun phrases, see Keenan and Westerstahl (1997) and the article **Quantifiers: Semantics**.

An example of a different nature is the following. (5) entails (6) (see Dowty, 1979):

- (5) John cools the soup.
- (6) The soup cools.

Here, one would like to have a description of the relation between the meaning of *cool* as a transitive causative verb and *cool* as an intransitive verb. And one would like to know why the implication holds for *boil* but not for *stir*.

A 'valid reasoning' is a generalization of entailment that involves more sentences than two. If someone accepts (7) and (8), they may conclude correctly (9):

- (7) John sings.
- (8) John is a man.
- (9) A man sings.

A more intricate example is (Groenendijk and Stokhof, 1982):

- (10) John knows whether Mary comes.
- (11) Mary comes.
- (12) John knows that Mary comes.

Here, one sees that there is a relation between the meaning of the *whether* clause and the *that* clause. It seems that the latter is weaker. However, the relation is more complicated: if one has the negated version (11a), then (12a) follows from (10) and (11a)

- (10) John knows whether Mary comes.
- (11a) Mary does not come.
- (12a) John knows that Mary does not come.

Hence the relation between the *whether* clause and the *that* clause depends on the factual situation.

A special case of entailment is 'synonymy.' Sentences are called synonymous just in case they entail

each other. For example, sentences (13) and (14) are synonymous:

(13) John or Mary comes.

(14) John comes or Mary comes.

Another example of synonymy is (Partee and Bach, 1981):

(15) Mary admires herself, and Sue does too.

(16) Mary admires herself, and Sue admires herself.

It may seem that the meaning of the *does too* clause can be found by the substitution of a phrase that occurs elsewhere in the sentence. This is, however, not always the case. It is, for instance, possible to conclude from (17) to (18).

(17) Mary believes that she is ill, and Sue does too.

(18) Mary believes that she is ill, and Sue believes that Mary is ill.

This illustrates the important phenomenon of ‘ambiguity.’ Sometimes a sentence can be understood in two or more ways corresponding with distinct consequences. From (17), one may conclude either that

(19) Sue believes that Mary is ill.

or that

(20) Sue believes that she (= Sue) is ill.

Another example of ambiguity is (e.g., Janssen, 1983, 1986b):

(21) Ten years ago, John met the president of the USA.

(22) The president of the USA is Bush.

On the one reading of (21), one may conclude that ten years ago John met Mr. Bush. On the other reading, this does not follow since John met the person who was president ten years ago. This ambiguity clearly concerns the functioning of tense operators.

The decision whether a sentence is ambiguous or vague is not always clear. Consider (23):

(23) Two girls are five sandwiches.

One may ask what counts as a source of semantic ambiguity; if it makes a difference whether they shared all five sandwiches; whether the sandwiches were distributed between the girls; or if it makes a difference whether they ate together or whether each girl ate on her own. Maybe the issue would be more exciting if the sentence was *Two girls shot five men* and one had to judge the girls. Verkuyl and Van der Does (1996) argue that (23) has one meaning, whereas they refer to other authors who argue for a fourfold or even ninefold ambiguity.

So far, only examples of relations between declarative sentences have been cited. Other types of sentences take part in ‘other semantic relations,’ as between the questions (24) and (25):

(24) Which girls came to the party?

(25) Did Mary come to the party?

In this example, there is nothing like the acceptance of a premise or of a conclusion. Nevertheless, there is a relation: every answer to the first implies an answer to the second question. The meanings assigned to these questions should account for this (Groenendijk and Stokhof, 1989).

Sequences of sentences likewise have logical properties. Consider (26) and (27).

(26) A man walks in the park. He whistles.

(27) A man in the park whistles.

This example puts a requirement on the treatment of texts. The meaning of the two sentences together should be equivalent to the meaning of the single sentence.

All examples discussed in this section have the common feature that they do not depend on specific details of the meanings of the words involved (except for logically crucial words such as *not* or *and*). It is, for example, possible to account for the inference from (1) *Mary is singing and dancing* to (2) *Mary is dancing* without describing the differences between *dancing*, *jogging*, and *walking*. Each of the examples can be replaced by another which exhibits the same pattern but uses other words. The examples illustrate that in Montague semantics one is mainly interested in structural aspects of the semantic relations between sentences, that is, in the systematic aspects of meaning. The formalization of what meanings are, does not need to go further than is required for an adequate account of these structural aspects.

The Compositional Approach

The most salient aspect of Montague semantics is the systematic way in which natural language expressions are connected with their respective meanings. This relation is characterized by the principle of compositionality of meaning. Such a principle can, in several formulations, be found in many disciplines that deal with semantics, such as linguistics, philosophy, and computer science. In philosophy of language, it has a long tradition and is often called ‘Frege’s principle.’ The version of this principle that describes the method of Montague grammar most explicitly is (Partee *et al.*, 1990: 318):

The meaning of a compound expression is a function of the meanings of its parts and of the syntactic rules by which they are combined.

The formulation of the principle contains several vague terms, and a proper application of the principle requires a more formal interpretation (for a discussion of the principle, its formalization, consequences, and its status, see **Compositionality, Aspects of**). The main points of its formalization are summarized here.

The syntax of the grammar consists of rules which express how new expressions can be formed from already available ones. The rules are, therefore, operations which act on inputs and yield an output. If an expression *E* is the output of the application of a rule *R*, then the inputs that form *E* are defined as being the parts of *E* in that derivation. The semantic component is organized in a parallel way: there are semantic rules that operate on input meanings and yield an output meaning. The crucial factor for obeying compositionality is that there is a strict correspondence between syntax and semantics. For each syntactic rule, there should be a corresponding semantic rule expressing the semantic effect of that syntactic rule. Compositionality is taken to be not an empirically verifiable property of natural language but a methodological principle: it constrains the organization of the grammar.

Consider example (28):

(28) Penguins do not fly.

A very simple grammar will be considered, naively as regards its meaning. For example, it will be assumed that the sentence says something about all penguins, whereas plurals without article usually have a more subtle meaning. The intention of this example is, however, only to illustrate the *method* of compositionality.

Suppose the grammar has as basic expressions the plural noun phrase *penguins* and the verb *fly*. A rule (say *R1*) forms the verb phrase *do not fly* from this verb. Furthermore, there is a rule (*R2*) combining a noun phrase with a verb phrase to form a sentence, by concatenating them and performing the required changes in the verb phrase for agreement and similar trimmings. Then sentence (28) has, according to this grammar, two parts: *penguins* and *do not fly*, and the latter phrase has in turn one part: *fly*.

This derivation might be represented in the form of a tree, as in **Figure 1**. Note that this tree does not

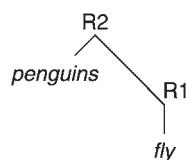


Figure 1 The derivation tree of 'Penguins do not fly.'

depict the constituent structure of the sentence; for example, there are no separate nodes for *do* and *not*. The tree shows how the sentence is formed; it is a construction tree or derivation tree. There is no *a priori* reason why the derivation would be identical to the constituent structure of the result (one might impose this as an additional requirement).

Of course, there might be good arguments for preferring a different grammar. Thus, one might construct (28) out of the positive sentence *Penguins fly*, or, alternatively, from *penguins*, *fly*, and *not*. In the former case the rule has one part, and in the latter case three parts. Compositionality as such provides no criterion for such issues. The best choice is probably to be discovered by considering more examples and larger fragments.

The principle of compositionality states that the meaning of the sentence is a function of the meanings of its parts, hence (according to the given grammar) of *penguins* and *do not fly*. Of course, the meaning of the latter is, in turn, a function of the meaning of its part *fly*. So, in the end, the meanings of the basic expressions are attained. Adopting for the moment a very simple conception of meaning that will be revised in the next section, one can take the meaning of *fly* to be the set of individuals who fly, and the meaning of *penguins* to be the set of individuals who are penguins.

According to the rules, the verb phrase *do not fly* has only *fly* as a part, and therefore the meaning of this verb phrase has to be formed out of the meaning of *fly*. So there has to be a semantic operation that expresses the meaning of negating a verb. For the moment, the meaning of *do not fly* is, in analogy of that of *fly*, the set of individuals who do not fly. The operation that forms this meaning from the meaning of *fly* is the formation of a set complement. In line with the above assumptions about meaning, one may say that the sentence means that the set of penguins is included in the set of nonflying individuals. This meaning is to be obtained from the meanings of its parts: from the set of penguins, and from the set of individuals who do not fly. This can indeed be done.

The situation is as follows. Two syntactic rules (*R1* and *R2*) are each accompanied by a semantic interpretation (*M1* and *M2*, respectively):

R1: negating a verb

M1: complement formation

R2: concatenating a noun phrase with a verb phrase, performing agreement changes

M2: set inclusion

In this section, the method of compositionality has been exemplified. The crucial aspect is the correspondence between syntax and semantics. One might change the concept of meaning used above, or change

the syntax used; but as long as the correspondence remains intact, the grammar can be seen as an instance of Montague grammar. This characterization of Montague's method is given in a formal mathematical terminology in his paper 'Universal grammar' (Montague, 1970b).

Interpretation in a Model

In Montague grammar, as well as in all other formal theories of semantics, the natural language expressions are interpreted in a class of abstract models. For example, a name like *John* is associated with an individual in such models, and an adjective like *brave* is a property. Each model is constructed out of a number of basic sets by means of standard constructions, and the result can be restricted by 'meaning postulates.' The most characteristic feature of the models in Montague grammar is the distinction made between the 'extension' of an expression and its 'intension,' a distinction that will be the topic of the next section. In the present section, the status of the model and its connection with natural language will be considered.

The model in which humans interpret natural languages has, of course, a certain resemblance to the real world, but it should not be conceived of as a model of the real world. There are two differences. First, in language, one speaks not only about the real world, past, present, and future, but also about situations that might be the case. Even though unicorns do not exist, one can speak about them, and the sentences used have semantic properties that should be dealt with. The model thus embraces much more than reality.

Second, as far as the model is connected with reality, it is a model of how natural language 'conceives' of it. This conception might be different from the real situation. Examples are mass nouns like *water* or *air*. In natural language, they are used in a different way from count nouns such as *chair* or *flower*. The mass noun *water* is used as if every part of a quantity of water is again water; as if it had no minimal parts. The same holds for *air*. Although in reality water has minimal molecules, and air is a mixture, the model will not reflect that fact (Pelletier and Schubert, 2003).

Although the model does not reflect reality, one can be interested in the relation which it has with reality. Examples are the relation between *blue* and the frequencies of light, or the number of exceptions accepted when it is said that *all ravens are black*. This kind of research is rare in Montague grammar, however, because it amounts to the analysis of specific words, whereas in Montague grammar one is mainly interested in the more structural aspects of the semantics of natural language.

The model, not being a model of reality, might be regarded as a model of how the human mind conceives reality. Although psychological reality is not one of the claims of Montague grammar (it is so in some other theories), the issue has received some attention (e.g., Dowty, 1979: ch. 8).

The connection between natural language and the model can be made in two ways. One method, the direct one, was followed in the section 'The Compositional Approach': for a word, some element (set, function) in the model is given as the meaning of that word, and for each syntactic rule a corresponding meaning operation is given. This method is used in Montague's first publication (Montague, 1970a), and in a few other publications as well (e.g., Keenan and Faltz, 1985). The other method is the indirect one: natural language is translated into some logical language, which is interpreted in a model. If this translation is compositional, and the interpretation of the logical language is compositional, then the combination of the two processes is a compositional process of meaning assignment. Care has to be taken that the logic is used as an auxiliary language only, so that this intermediate language can in principle be eliminated. This implies that every operation that is performed in the logic should have an interpretation in the model. This indirect method is the standard method in Montague grammar; usually, (variants of) intensional logic are used as the intermediate language.

Extension and Intension

An important decision concerns the question of how to model meaning. Previously, the meaning of *penguins* was defined as the set of penguins, and of *fly* as the set of flying individuals. Applying the same approach to *the president of the USA* [example (22)], would yield as meaning the individual Bush, and applied to *unicorn* (assuming there are none) the empty set. This approach would, however, give results that are intuitively incorrect. It would have the consequence that in case neither unicorns nor centaurs exist, the meaning of *unicorn* would be equal to the meaning of *centaur*. As for *the president of the USA*, it would have the undesirable consequence that its meaning changes after each election. Examples like these have led to the distinction between two kinds of interpretation: extension and intension.

At the time of writing, the president of the USA is Mr Bush, but at other moments in time a different person will be president, and in another course of events Mr Bush could have lost the election and Mr Kerry would be president. The model in which natural language is interpreted has components dealing with this. It has a collection of time points for

dealing with changes in the course of time, and it has a collection of so-called 'possible worlds.' These possible worlds represent the possibility that Kerry has won. They also represent the possibility that unicorns exist. Intensions are functions with possible worlds and time points as domain. The intension of *the president of the USA* is a function from time points and possible worlds that yields an individual (the president at that moment in that possible world), and the intension of *unicorn* is the function that yields for each possible world and time point a set of individuals (the unicorns). The extension of an expression is the value of the intension function with respect to a given world and time point, for example, the moment now in the actual world. Then the extension of *the president* is the president now (Mr Bush), and the corresponding extension of *unicorn* is the actual set of unicorns. The extension of a sentence is traditionally identified with its truth value. Thus, the extension of *John kisses Mary* is true just in case John kisses Mary. The intension is the function which says, for each possible world, in which moments John kisses Mary.

Since there are possible worlds in which there are unicorns but no centaurs, the words *unicorn* and *centaur* have different intensions. As a consequence, sentences (29) and (30) will have different intensions too.

(29) John seeks a unicorn.

(30) John seeks a centaur.

Thus, using intensions, the nonsynonymy of (29) and (30) can be accounted for. For this purpose, no further information concerning relations between different possible worlds is needed, nor any information concerning relations between time points. This holds for all examples mentioned in the section 'Aims,' and therefore the set of possible worlds and the set of time points are usually introduced as just sets without further specification. This is, however, not always sufficient.

If one is interested in tense phenomena in natural language, then more has to be said about the moments of time, for example that they are linearly ordered, and whether they are indivisible points or intervals with a duration. If one is interested in causatives (*John broke the glass*) or counterfactuals (*If Mary did not come, John would fly*), then the set of possible worlds needs more structure. For dealing with these phenomena, it is crucial to know how the world was just before the breaking of the glass, or which world resembles the present one except for the coming of Mary.

The above discussion shows that the formalization of the intuitive notion of meaning as intension is much better than as extension. However, intensions

only deal with those aspects of meaning that have to do with truth and denotation, and neglect aspects such as style, new information versus old information, etc. Therefore, they can only be regarded as a restricted approximation of meaning. Even accepting this limitation, however, intensions are still not completely satisfactory. At important shortcoming concerns tautologies. Consider (31):

(31) Bill comes or does not come.

The intension of this sentence is the function that always yields 'true.' Hence (32) and (33) have the same intension.

(32) John comes.

(33) John comes and Bill comes or does not come.

This causes problems with embedded clauses. Sentences (34) and (35) will have the same intension, whereas they should not be equivalent since Mary's beliefs in (34) do not concern Bill; she may not even know about his existence:

(34) Mary believes that John comes.

(35) Mary believes that John comes and that Bill comes or does not come.

The conclusion is that intensions are not fine-grained enough to distinguish the meanings of sentences like (34) and (35). Several improvements have been proposed, such as structured intensions (Lewis, 1970), and an approach based on partial functions (Muskens, 1989).

A Small Fragment

This section presents as examples three sentences and their treatment. The treatment given in Montague (1973) will be followed, except for one detail (see next section). In the course of the presentation, some important features of intensional logic will be explained.

The sentences are:

(36a) John walks.

(37a) A man walks.

(38a) Every man walks.

These sentences are very simple, and it is easy to present their interpretation in traditional predicate logic:

(36b) $walk(john)$

(37b) $\exists x[man(x) \wedge walk(x)]$

(38b) $\forall x[man(x) \rightarrow walk(x)]$

One immediately sees that these three sentences are syntactically much alike (a subject and a verb), but the formulas in predicate logic are rather different: two with a quantifier (different ones!), and in each case another main operator (\rightarrow , \wedge , and predicate application). What makes it interesting to present these examples here is not the meaning assigned to them, but the fact that these very different meanings can be derived compositionally from the corresponding sentences.

All three sentences are built from a singular noun phrase and the verb *walks*, and for this reason one can design a single syntactic rule that combines a noun phrase and a verb. Since in Montague grammar there is a correspondence between syntactic rules and semantic rules, one also has to design a rule that combines the meaning of the verb with the meaning of the noun phrase. This in turn requires that there be meanings for verbs and for noun phrases. These meanings will be discussed first, then the semantic rule will be designed.

As explained in the previous section, predicates are given intensions as meanings: functions from possible worlds and moments of time to sets of individuals. Thus, the intension, or meaning, of the verb *walk* is a function from possible worlds and moments of time to sets: for each possible world and each moment of time, there is a set (possibly empty) of individuals who walk. Such an intension is called a 'property.' For the noun phrases, it is more difficult to select a format that allows these meanings to be rendered uniformly. In order to keep the calculus going in a maximally uniform way, all noun phrase meanings should preferably be of the same type. This requirement is easily seen to be nontrivial, as, for example, the expression *every man* extends over sets of possibly many individuals, where *John* seems to refer to one individual only.

Montague proposed (1973) to model the meaning of a noun phrase as a set ('bundle') of properties. An individual is, in this approach, characterized by the set of all its properties. This is possible since no two individuals share all their properties at the same time, if only because the two cannot be at the same place at the same time. This approach permits generalization over all noun phrases referring to (sets of) individuals (see also Lewis, 1970).

Meanings of linguistic expressions (functions from world-time pairs to specific world extensions) are denoted by formulas in the language of intensional logic. The meaning of *John* will be denoted by:

$$(39) \lambda P[\forall P(\text{john})]$$

To explain this formula, the variable

$$(40) P$$

is a variable over properties of individuals: P may be replaced, for example, by the property 'walking.' The expression

$$(41) \forall P$$

denotes the extension of the property expressed by any predicate P in any given world at a given moment of time. Thus, for the actual world now.

$$(42) \forall P(\text{john})$$

is true if the predicate $\forall P$ holds for John now in this world, and false otherwise. Any property denoted by P can be abstracted by means of the lambda operator: λP . Lambda abstraction of (42) gives:

$$(43) \lambda P[\forall P(\text{john})]$$

which is the same as (39) above. This expression denotes a function which says, for any property α , given a world-time pair $\langle w, t \rangle$, whether John belongs to the extension of that property in $\langle w, t \rangle$. Let this function be called X_j . Some properties of this function will now be investigated.

According to the definition of X_j :

$$\begin{aligned} X_j(\alpha) \text{ is true if and only if } \alpha \text{ now holds for} \\ \text{the individual John, i.e., } \forall \alpha(\text{john}) \text{ is true,} \\ \text{and false otherwise} \end{aligned} \quad (I)$$

X_j is, therefore, the characteristic function of the set of properties that can be predicated of John. As usual in logic, this characteristic function can be identified with the set of properties that John has. X_j can therefore be seen as a formalization of the idea that the meaning of *John* is a set of properties.

This function X_j can be evaluated with as argument the property of being a man, that is, the function that yields for each index (i.e., for each world and time) the extension of the predicate *man*. The notation for this argument is:

$$(44) \wedge \text{man}$$

The symbol \wedge translates the predicate *man* into the language of intensional logic, where $\wedge \text{man}$ denotes the intension associated with *man*. X_j is now applied to this argument to obtain, using result (I).

$$\begin{aligned} X_j(\wedge \text{man}) \text{ is true if and only if } \wedge \text{man} \text{ now} \\ \text{holds for the individual John, i.e., if and only} \\ \text{if } \forall \wedge \text{man}(\text{john}) \text{ is true, and false otherwise} \end{aligned} \quad (II)$$

In the expression $X_j(\wedge \text{man})$, X_j can be replaced by its original definition (viz. (43)). Then, (45) is obtained.

$$(45) \lambda P[\forall P(\text{john})](\wedge \text{man})$$

Result (II) now states that this is equivalent with

$$(46) \quad \forall^{\wedge} \text{man}(\text{john}).$$

Thus, it has been shown, using semantic considerations, that it is allowed to substitute the argument $\wedge \text{man}$ for the variable P , an operation known as ‘lambda conversion.’ According to the definitions given with (41) and (44), $\forall^{\wedge} \text{man}$ denotes the present value of the property of being a man. Hence (46) and (45) are equivalent with (47)

$$(47) \quad \text{man}(\text{john})$$

Summarizing, the variable P in (45) has been replaced with the property $\wedge \text{man}$, which is in the range of P , and $\forall^{\wedge} \text{man}$ has been reduced to just man . As the operations concerned are admissible independently of the particular choice of the predicate man , the procedure can be generalized to all predicates of the same class (type).

To revert to the treatment of the simple sentence *John walks*, the syntax has, as stipulated earlier, a rule that combines a noun phrase with a verb to form a sentence. What is still needed is a semantic operation that matches the syntactic rule with its semantic consequences. The operation that does this is so-called ‘function application’: one of the two syntactic constituents acts as a function, while the other acts as argument (input). In this case, the verb meaning is allowed to act as the argument, and the noun phrase meaning as the function. The result of function application, in this case, is a function from world–time pairs to truth-values, or the set of world–time pairs where John walks.

According to the rule just given, the meaning of *John walks* is found by application of the meaning of *John*, that is (43), to the meaning of *walk*. This is denoted by:

$$(48) \quad \lambda P[\forall P(\text{john})](\wedge \text{walk})$$

This, as seen above, can be reduced, in two steps, to:

$$(49) \quad \text{walk}(\text{john})$$

which now gives the meaning representation aimed at.

For the other sentences, one proceeds analogously. The noun phrase *a man* is likewise translated as a set of properties:

$$(50) \quad \lambda P[\exists x[\text{man}(x) \wedge \forall P(x)]]$$

This denotes the characteristic function of the set of properties such that for each property in the set there is a man that has that property. The sentence *A man walks* is then represented as:

$$(51) \quad \lambda P[\exists x[\text{man}(x) \wedge \forall P(x)]](\wedge \text{walk})$$

which reduces to:

$$(52) \quad \exists x[\text{man}(x) \wedge \text{walk}(x)]$$

Analogously, every man walks is represented as.

$$(53) \quad \lambda P[\forall x[\text{man}(x) \rightarrow \forall P(\text{john})]](\wedge \text{walk})$$

or equivalently

$$(54) \quad \forall x[\text{man}(x) \rightarrow \text{walk}(x)]$$

This treatment illustrates some of the power of lambda abstraction and lambda conversion. The meaning of the verb is ‘plugged into’ the meaning of the noun phrase in the right position. Lambda calculus is frequently used in Montague grammar. Without the lambda operator, it would be impossible to maintain compositionality. Impressed by the power of lambdas, Barbara Partee once said: ‘Lambdas really changed my life.’ What has been, in the end, obtained as meaning representations for the three sentences discussed is nothing more than the formulas usually associated with them in elementary logic courses. There, however, they are found on intuitive grounds, whereas in Montague grammar they are the result of a formal system which relates syntax and semantics in a systematic way.

Some PTQ Phenomena

Montague worked out his ideas in a number of papers, the most influential of which is ‘The proper treatment of quantification in ordinary English,’ henceforth PTQ (Montague, 1973). This article deals with some semantic phenomena, all connected with quantifiers. Three such phenomena will be discussed here: identity and scope ambiguities (both presented here because they have been the subject of a great deal of discussion since the publication of PTQ), and the ‘de dicto–de re’ ambiguity, central to Montague grammar, and the origin of its trade mark, the unicorn.

The first phenomenon concerns problems with identity. Consider:

$$(55) \quad \text{The price of one barrel is \$1.00.}$$

$$(56) \quad \text{The price of one barrel is rising.}$$

$$(57) \quad \$1.00 \text{ is rising.}$$

It is obvious that (57) must not be allowed to follow logically from (55) and (56), as \$1.00 will remain \$1.00 and will neither rise nor fall. The same phenomenon occurs with temperatures, names, telephone numbers, percentages, and in general with all nouns which may denote values. The PTQ treatment of such cases is as follows. Prices, numbers, etc. are treated as basic entities, just as persons and objects. The expression *the price of one barrel* is semantically

a function that assigns to each world–time index a particular price. Such a function is called an ‘individual concept.’ In (55) an assertion is made about the present value of this function, but in (56) an assertion is made about a property of the function. Thus, the expression *the price of one barrel* is considered ambiguous between a value reading and a function reading. The difference in readings blocks the false conclusion (57).

In spite of criticisms, mainly related to the treatment of such nouns as basic entities (e.g., Bennett, 1976), the notion of individual concepts also seems useful to account for cases like the following (Janssen, 1984).

- (58) The treasurer of the charity organization is the chairman of the hospital board.
- (59) The treasurer of the charity organization has resigned.
- (60) The chairman of the hospital board has resigned.

Here, substitution *salva veritate* of terms under an identity statement seems to be running into difficulty. Again, one can say that (58) is a statement about the identity of the values of two different functions at a given world–time index, whereas (59) is a statement about the function (in Frege’s terms, about the *Wertverlauf*). Hence the nonvalidity of (60) as a conclusion is explained.

The second phenomenon is scope ambiguity. Consider

- (61) Every man loves a woman.

In PTQ, this sentence is considered to be ambiguous (though many linguists would disagree; see below). In the one reading, one particular woman is loved by all men, and in the other every man has at least one woman whom he loves. The first reading is given in (62), and is called the specific reading, or the wide scope reading for *a woman*. The second reading is given in (63); it is called the ‘narrow scope reading.’

- (62) $\exists x[\text{woman}(x) \wedge \forall y[\text{man}(y) \rightarrow \text{love}(y, x)]]$
- (63) $\forall y[\text{man}(y) \rightarrow [\exists x \text{ woman}(x) \wedge \text{love}(y, x)]]$

Many linguists consider (61) to be unambiguous. Well-known is the principle which states that the most plausible reading is given by the surface order of the quantified NPs (and of the negation). Following this principle, (61) has only one reading, viz. (63). Note that the reading expressed by (62) is a special case of (63). The principle has not remained unchallenged. Witness (64), where the most plausible reading is not given by the surface order.

- At the finish, a medal is available for all participants in the race.

There are other linguistic theories which also assign one reading to (61). But whether (61), the PTQ example, really is ambiguous is less relevant as long as there are sentences that do show scope ambiguities, which seems beyond doubt. For instance, previously an example involving tense was given (*Ten years ago, John met the president*). The machinery of PTQ for dealing with scope ambiguities is presented below.

Since the scope ambiguity does not seem to have a lexical source, the principle of compositionality of meaning requires the construction of two different derivations for the two readings. In PTQ, this is done as follows. First, the basic sentence (65) is formed, in which neither of the two noun phrases occurs but which contains indexed variables (he_n) instead:

- (65) $he_1 \text{ loves } he_2$.

Next, noun phrases are substituted for the variables. This can be done in two different orders. The noun phrase that is substituted last gets the widest reading. Thus, the specific reading (wide scope for *a woman*) is obtained from (65) by first forming (66).

- (66) Every man loves he_2 .

and then

- (67) Every man loves a woman.

These rules are called ‘quantification rules.’ The corresponding semantic rule leads to an interpretation equivalent to that of (62). For the other reading, the reverse order of quantifier substitution is used.

These quantification rules have met with some resistance from linguistic quarters, where they are looked upon as unusual creatures. Other solutions have been attempted, where the disambiguation is not done in the syntax. Cooper (1979) deals with scope phenomena in a separate component, a ‘storage’ mechanism. Hendriks (1993) uses rules which change the meaning of a noun phrase from a narrow scope reading to a wider scope reading.

The third phenomenon is the ambiguity of ‘de dicto’ versus ‘de re’ readings. To see the difference, first consider (68) and (69):

- (68) John finds a unicorn.
- (69) There is a unicorn that John finds.

Sentence (69) follows from (68). Yet (71) does not follow from (70):

- (70) John seeks a unicorn.
- (71) There is a unicorn that John seeks.

In fact, (70) is ambiguous between a specific reading in which there is a specific unicorn that John seeks,

and an intensional reading where John is said to engage in the activity of seeking a unicorn and nothing is implied about the real existence of such animals. The latter reading is usually called the ‘de dicto’ reading (‘de dicto’: Latin for ‘about what is said’). The former reading is the ‘de re’ reading (‘de re’: Latin for ‘about the thing’). The ambiguity is accounted for, in principle, in the following way. In the ‘de re’ reading, a relation is asserted to hold between two individuals. The ‘de dicto’ reading establishes a relation between John and the set of properties of a unicorn (i.e., the interpretation of the noun phrase *a unicorn*). Whether this way of analyzing and accounting for the ‘de dicto’ reading is satisfactory in all respects is a question still widely debated (see e.g., Gamut, 1991).

The two readings of *John seeks a unicorn* are obtained in the following way. For the ‘de re’ reading, a quantification rule introduces the expression *a unicorn*. And if the expression *a unicorn* is introduced directly, that is, without the intervention of a quantification rule, the result is the ‘de dicto’ reading. In principle, this method is analogous to the method used to express scope differences. One may, in fact, regard the distinction between ‘de dicto’ and ‘de re’ readings as a distinction due to scope differences of the intensional operator *seek*.

Developments

Montague (1973) (henceforth PTQ) signaled the start of a great amount of semantic research. Nowadays there are many linguistic theories that deal in a formal way with semantics, and certainly not all of them obey Montague’s compositional framework, but in one respect they are all indebted to him: lambdas are used frequently (see section ‘A Small Fragment’ for the power of lambdas). This section reviews some of the developments that have emanated since PTQ and remain more or less within his compositional framework.

In the first period one kept close to the style of PTQ: one widened the original fragment with new phenomena. A characteristic example is the title ‘A variation and extension of a Montague fragment of English’ (Bennett, 1976). In such articles one finds treatments of relative clauses, passives, scope, control, numbers, plurals, generics, complement structures, and deictic pronouns: most deal with English, some with other languages. For references to these early works see Partee (1976), Gamut (1991), and Dowty *et al.* (1981). Nowadays the habit of providing explicit fragments is abandoned: one focuses on some semantic phenomenon and provides rules for that phenomenon only.

The first broadening of the scope of Montague semantics was toward lexical semantics. In PTQ the words are regarded as basic and remain unanalyzed except for a few logically interesting words such as the verb *be*: often, however, it is linguistically interesting to carry the analysis of words further. Examples (5) and (6) featured the word *cool*, which can be an adjective, an intransitive verb phrase, and a causative verb, with related meanings. Such lexical analyses are given in Dowty (1979).

Montague used a primitive kind of syntax. On the one hand, this appears from the fact that the syntactic rules are not subject to any formal restriction; on the other hand, from the absence of grammatical features and a well-motivated constituent structure and grammatical functions. Soon after the publication of PTQ, proposals were made to incorporate syntactic know-how as developed in transformational grammar into Montague grammar (see, e.g., Partee, 1975 and Dowty, 1982). Partee’s ideas for combining Montague grammar with transformational grammar are worked out in the large grammar used in the machine translation project Rosetta (Rosetta, 1994).

Since the earliest years of Montague semantics there have been attempts to make the syntactic rules as restricted as possible by using only concatenation rules. As a consequence, the constituent structure of the natural language expression becomes isomorphic to its compositional analysis. Because of this feature, such approaches are said to obey ‘surface compositionality.’ The approach leads to (variants of) categorical grammar. An early advocate of this approach was Bach (see, e.g., Oehrle *et al.*, 1988); a recent monograph within the framework of Montague is Morill (1994). Surface compositionality requires a treatment of scope phenomena different than Montague’s. An important idea is the use of type shifting rules: such rules make the relation between a word and its semantic type less rigid than in PTQ, and therefore this approach is called ‘flexible Montague grammar.’ The method is useful for several phenomena (including scope); see Partee and Rooth (1983), Hendriks (1993), and Groenendijk and Stokhof (1989).

As explained previously, noun phrases are treated in PTQ as sets of properties. Consequently the determiner ‘every’ is a function from sets to sets: it can be conceived of as a relation between two sets (e.g., in ‘Every man dances’ as a relation between ‘man’ and ‘dance’). This abstract analysis of noun phrases opened the way to a new field: the theory of ‘generalized quantifiers,’ in which many phenomena concerning noun phrases are studied. The pioneering article is Barwise and Cooper (1981); monographs

are Keenan and Faltz (1985) and Winter (2001), (*see Quantifiers: Semantics*).

The largest linguistic units analyzed in PTQ are sentences. The task of extending Montague's method to discourses looked unrewarding at first, particularly in view of cross-sentential anaphora phenomena. The first step from sentence to discourse (Kamp, 1981) was, on certain details, not compositional. Later a compositional treatment was given (Groenendijk and Stokhof, 1991). The key feature of their approach was that the meaning of a sentence is not a truth value, but its information change potential: a function that changes the information state. This led to dynamic semantics, an approach in which several other 'information' phenomena can be analyzed (*see Dynamic Semantics*).

Also, the ontology used by Montague has been challenged. His ontology was based upon a hierarchy of typed sets, but several phenomena suggest a change. Mass nouns seem to ask for sets that do not have elements (Pelletier and Schubert, 2004). Nominalizations (as in *I love loving*) seem to ask for an alternative in which functions can have themselves as arguments; this leads to property theory (see, e.g., Chierchia, 1982). Perception verbs challenge the basic role of possible worlds. They are abandoned by Barwise and Perry (1983), who replace them by 'situations,' but a classical model with partial functions can also be used (Muskens, 1989). PTQ deals with only one type of language utterances, declarative sentences, and one may think that other utterances require other types. But questions can be treated without a new ontology, as is shown by the work on the semantics of questions and wh-complements (Groenendijk and Stokhof, 1982, 1989).

Further Reading

A well-known introduction to Montague grammar (especially written for linguists) is Dowty *et al.* (1981). An introduction with much attention paid to logical and philosophical motivation is Gamut (1991). An introductory article on 'Universal grammar' (Montague, 1970b) is Halvorsen and Ladusaw (1979), its mathematical background is investigated in Janssen (1986a) and discussed in Janssen (1997). Publications on Montague grammar often appear in *Linguistics and Philosophy*. Many references are given in Partee and Hendriks (1997), recent collections of papers about Montague grammar are Partee (2004), and Portner and Partee (2002).

See also: Compositionality: Semantic Aspects; Dynamic Semantics; Extensionality and Intensionality; Formal Semantics; Quantifiers: Semantics.

Bibliography

- Barwise J & Cooper R (1981). 'Generalized quantifiers and natural language.' *Linguistics and Philosophy* 4, 159–219.
- Barwise J & Perry J (1983). *Situations and attitudes*. Cambridge, MA: Bradford Books.
- Bennett M (1976). 'A variation and extension of a Montague fragment of English.' In Partee (ed.), 119–163.
- Chierchia G (1982). 'Nominalization and Montague grammar, semantics without types for natural languages.' *Linguistics and Philosophy* 5, 303–354.
- Cooper R H (1979). 'The interpretation of pronouns.' In Heny F & Schnelle H (eds.) *Selections from the third Groningen round table. Syntax and Semantics* 10. New York: Academic Press. 61–92.
- Dowty D R (1979). *Word meaning and Montague grammar*. Synthese Language Library 7. Dordrecht: Reidel.
- Dowty D (1982). 'Grammatical relations and Montague grammar.' In Jacobson P & Pullum G K (eds.) *The nature of syntactic representation*. Dordrecht: Reidel. 79–130.
- Dowty D R, Wall R E & Peters S (1981). *Introduction to Montague semantics*. Synthese language library 11. Dordrecht: Reidel.
- Gamut L T F (1991). *Logic, language and meaning, vol II: Intensional logic and logical grammar*. Chicago, London: The University of Chicago Press.
- Groenendijk J & Stokhof M (1982). 'The syntax and semantics of wh-complements.' *Linguistics and Philosophy* 5, 175–233.
- Groenendijk J & Stokhof M (1989). 'Type-shifting rules and the semantics of interrogatives.' In Chierchia G, Partee B H & Turner R (eds.) *Properties, types and meaning, vol II: Semantics issues*. Dordrecht: Kluwer. 21–68. [Reprinted in Portner & Partee (2002). 421–456.]
- Groenendijk J & Stokhof M (1991). 'Dynamic predicate logic.' *Linguistics and Philosophy* 14, 39–100.
- Groenendijk J A G, Janssen T M V & Stokhof M B J (eds.) (1981). *Formal methods in the study of language. Proceedings of the third Amsterdam colloquium*. MC-Tracts 135 and 136, Amsterdam: CWI.
- Halvorsen P K & Ladusaw W A (1979). 'Montague's "Universal Grammar". An introduction for the linguist.' *Linguistics and Philosophy* 3, 185–223.
- Hendriks H (1993). *Studied flexibility. Categories and types in syntax and semantics*. ILLC dissertation series 1993–5. Amsterdam: Institute for Logic Language and Computation, University of Amsterdam.
- Janssen T M V (1983). 'Scope ambiguities of tense, aspect and negation.' In Heny F & Richards B (eds.) *Syntactic categories, auxiliaries and related puzzles*. Synthese Language Library 20. Dordrecht: Reidel. 55–99.
- Janssen T M V (1984). 'Individual concepts are useful.' In Landman F & Veltman F (eds.) *Varieties of formal semantics*. Grass 3. Dordrecht: Foris. 171–192.
- Janssen T M V (1986a). *Foundations and applications of Montague grammar. Part 1 Foundations, logic, computer science*. CWI Tract 19. Amsterdam: Centre for Mathematics and Computer Science.
- Janssen T M V (1986b). *Foundations and applications of Montague grammar, Part 2: Applications to natural*

- language. CWI Tract 28. Amsterdam: Centre for Mathematics and Computer Science.
- Janssen T M V (1997). 'Compositionality.' In van Benthem & ter Meulen (eds.), 417–473.
- Kamp H (1981). 'A theory of truth and semantic representation.' In Groenendijk *et al.* (eds.), 277–322. [Reprinted in Portner & Partee (2002). 189–222.]
- Keenan E L & Faltz L M (1985). *Boolean semantics for natural language*. Dordrecht: Kluwer.
- Keenan E L & Westerstahl D (1997). 'Generalized quantifiers in linguistics and logic.' In van Benthem & ter Meulen (eds.), 837–893.
- Lewis D (1970). 'General semantics.' *Synthese* 22, 18–67. [Reprinted in Davidson D & Harman G (1970). *Semantics of natural language*. Synthese Language Library 40. Dordrecht: Reidel. 169–248; Partee (1976). 1–50.]
- Montague R (1970a). 'English as a formal language.' In Visentini *et al.* (eds.) *Linguaggi nella societa et nella tecnica*. Edizioni di communita (distributed by the Olivetti Corporation, Milan). [Reprinted in Thomason (1974). 188–221.]
- Montague R (1970b). 'Universal grammar.' *Theoria* 36, 373–398. [Reprinted in Thomason (1974). 222–246.]
- Montague R (1973). 'The proper treatment of quantification in ordinary English.' In Hintikka K J J, Moravcsik J M E & Suppes P (eds.) *Approaches to natural language*. Synthese Library 49. Dordrecht: Reidel. 221–242. [Reprinted in Thomason (1974). 247–270 and in Portner & Partee (2002). 17–34.]
- Morrill G (1994). *Type logical grammar, categorial logic of signs*. Dordrecht: Kluwer.
- Muskens R (1989). 'Going partial in Montague grammar.' In Bartsch R, van Benthem J & Van Emde Boas P (eds.) *Semantics and contextual expression*. Grass-series 11. Dordrecht: Foris. 175–220.
- Oehrle R T, Bach E & Wheeler D (1988). *Categorial grammars and natural language structures*. Dordrecht: Kluwer.
- Partee B (1975). 'Montague grammar and transformational grammar.' *Linguistic Inquiry* 6, 203–300.
- Partee B H (ed.) (1976). *Montague grammar*. New York: Academic Press.
- Partee B H (2004). *Compositionality in formal semantics*. Oxford: Blackwell.
- Partee B & Bach E (1981). 'Quantification, pronouns and VP-anaphora.' In Groenendijk *et al.* (eds.). [Reprinted in Groenendijk, Janssen & Stokhof (eds.). *Truth, interpretation, and information. Selected papers from the third Amsterdam colloquium*. Grass-series 2, Dordrecht: Foris. 99–130.]
- Partee B H & Hendriks H L W (1997). 'Montague Grammar.' In van Benthem & ter Meulen (eds.), 7–91.
- Partee B & Rooth M (1983). 'Generalized conjunction and type ambiguity.' In Bauerle R, Schwarze Ch & von Stechow A (eds.) *Meaning, use and interpretation of language*. Berlin: De Gruyter. 361–384. [Reprinted in Portner & Partee (2002). 334–356.]
- Partee B H, ter Meulen A & Wall R E (1990). *Mathematical methods in linguistics*. Dordrecht: Kluwer.
- Pelletier J F & Schubert L K (2003). 'Mass expressions.' In Gabbay D M & Guenther F (eds.) *Handbook of philosophical logic*, 2nd edn. Kluwer. 249–335.
- Portner P & Partee B H (2002). *Formal semantics. The essential readings*. Oxford: Blackwell.
- Rosetta M T (1994). *Compositional translation*. Dordrecht: Kluwer.
- Thomason R H (ed.) (1974). *Formal philosophy. Selected papers of Richard Montague*. New Haven: Yale University Press.
- Van Benthem J & ter Meulen A (eds.) (1997). *Handbook of logic and language*. Amsterdam & Cambridge: Elsevier & MIT Press.
- Van der Does J & Verkuyl H (1991). 'The semantics of plural noun phrases.' In van der Does J & van Eyck J (eds.) *Generalized quantifier theory and applications*. Dutch network for Logic, Language and Information. 403–441.
- Winter Y (2001). *Flexibility principles in Boolean semantics*. Cambridge, MA: MIT Press.

Mood, Clause Types, and Illocutionary Force

K Allan, Monash University, Victoria, Australia

© 2006 Elsevier Ltd. All rights reserved.

Sadock and Zwicky (1985) surveyed 35 languages representing a wide range of language families and linguistic areas. They found that every language distinguishes at least three **clause types**: declaratives (1), interrogatives (2), and imperatives (3). It is surely no accident that these are the clause types orthographically marked by '.', '?', and (often) '!', respectively. Clause type is formally defined by morphosyntax, lexis, and prosody.

- (1) He just wouldn't stop whining. Bill was shot by Max. Bill, Max shot. Bill died. Max gave the Salvos a donation. *It's snowing* is a declarative sentence. Mortgage rates down at last. Dunno who did it.
- (2) Will you stop whining? Care to come along? Which came first? The chicken or the chicken nuggets? Why bother? After all, who wants to crash a convivial high by spending an hour at the sink? I shouldn't have worn those heels today, should I? He's not bad, huh? Please may I be excused? You were where on the night of eighteenth May?

- (3) Stop your bloody whining! Move it, Grandma. Be warned! Get yourself checked by someone qualified. You give me that ball, you! Nobody move! Let the Games begin. Let me help. Let's eat. Have a good day!

The western classical tradition in linguistics identifies three **moods**: indicative, subjunctive, imperative – which partially correspond to clause types. Protagoras (490–420 BCE, Diogenes Laertius, Book IX) splits the indicative into interrogative [erōtēsis] and declarative [apokrisis]. Palmer (2001) recognizes only two moods: indicative and subjunctive, which he compares with the realis-irrealis division proposed for some Native American languages (Chafe, 1995; Mithun, 1995) and Papuan languages (Roberts, 1990); and also with the assertive/non-assertive division proposed by Quirk, Greenbaum *et al.* (1985), Wierzbicka (1988), and Lunn (1995). Cross linguistically, there is a cline between indicative/realis/assertive at one pole and subjunctive/hypothetical/irrealis/uncertain at the other (Givón, 1994; Wehmeier, 2004).

The semantics of a clause type is its **primary illocution** (PI). All languages have a **declarative** as the most frequent and least marked clause type; here its PI is symbolized **T**, because its token potentially bears a truth value when used by a speaker in an utterance in a certain context (*see Type versus Token*). The declarative is typically realis (though it may include a hypothetical). At the opposite extreme is the **PI H**, characteristic of hypotheticals, counterfactuals, intentionals, traditional subjunctives, and other irrealities. Like the subjunctive (Givón, 1994: 277; Jespersen, 1958: 318), hypotheticals often have no unique morphology. Between the extremes of **T** and **H** are strung interrogatives, imperatives, negatives, futures, and habituales which may align with **T** or with **H**. For instance, 'Jim is not here' describes an event that can be seen as a factual statement having a truth value; and therefore aligned with **T**; alternatively, it is a counterfact aligned with **H**. A question 'Who is that?' has no truth value (*It is true that who is that), thus it can align with **H**; but in the Western classical tradition it is indicative and aligned with **T**. In the space here, it is unrealistic to force all languages into a universal mold any more specific than that already described; so the focus is on English.

Coincidence between clause type and illocutionary force may have been recognized by Protagoras, certainly by Apollonius Dyscolus (100 CE); cf. Householder (1981); Sanctius (1585); Lancelot (1644); Lane (1700); Whitney (1888); Sadock (1974); Van Valin and LaPolla (1997); and Huddleston and Pullum (2002).

How does illocutionary force enter the picture?

- Linguistics is concerned (*inter alia*) with utterances in which Speaker uses a language expression and thereby performs a locutionary act; to do so, Speaker uses an identifiable sentence or clause fragment from a language *L* (sentence type, intonation type, etc.). The resulting token is a locution.
- Producing the locution demands that Speaker has knowledge of the morphosyntax, lexicon, semantics, and phonology of the language used. Recognizing the locution demands that Hearer has comparable knowledge.
- A locution will normally bear more than one illocutionary force (e.g., 'I'll see you at 10' is a statement, prediction, (potentially) commissive); yet, it will usually have only one message to convey (depending on context, perhaps a promise or a threat). The illocutionary force that carries this message is called the illocutionary point of the utterance.
- The form of the locution in the utterance must be the starting point for Hearer's interpretation of the utterance meaning. Thus, Hearer must seek to recognize (a) the locution, including the sense of each clause; (b) what Speaker intends the locution to refer to; and (c) the illocutionary forces within the utterance that give rise to its illocutionary point. Hearer hears the locution, recognizes its sense, looks to the context to figure out the apparent reference, and then seeks to infer Speaker's illocutionary intention.
- Speaker's illocutionary intention is to have Hearer recognize the illocutionary point of Speaker's utterance in order to achieve a particular perlocutionary effect – the behavioral and/or cognitive and/or emotional response to the illocutionary forces in the utterance. Speaker's perlocutionary act is an act of achieving a particular perlocutionary effect on Hearer as a result of Hearer recognizing (what s/he takes to be) the locution and illocutionary forces in Speaker's utterance.
- Recognizing the clause type is the first step toward discovering the illocutionary point via the PI. It is important to reaffirm that the PI is the semantic property of the clause type (details are given shortly). The PI of 'I'll see you at 10' is a statement (with its potential truth value); the statement makes a prediction about the Speaker's future behavior which, entering the pragmatic domain, typically constitutes a commissive; in turn, this commissive may be a promise or a threat depending on several contextual factors. The PI is one semantic basis for the pragmatic inferencing that leads to the illocutionary point.

What criteria identify clause type? Word order distinguishes declarative 'He is tall' from interrogative 'Is he tall?' Morphosyntactic marking distinguishes

the hypothetical backshifted modal in the temporally non-past ‘**Could** you mail this letter for me?’ from the indicative ‘**Can** you mail this letter for me?’, and these two interrogatives from the declarative ‘I can certainly mail it for you.’ Intonation distinguishes declarative from interrogative versions of ‘John’s gone to New York(./?)’ Similarly, in other languages: In Navajo, the imperative has the same morphosyntax as the declarative, but is prosodically distinct; in Korean, declarative, interrogative, imperative, and propositive (‘let’s’ constructions) are formally identical in intimate and polite speech levels, but are distinguished by prosody. In addition to prosodic marking, the interrogative ‘John’s gone where?’ is lexically marked by the *wh*- word at the questioned location. Lexical marking is significant in identifying interrogatives in subordinate clauses: ‘She asked **how/ whether / if** John had traveled to New York.’

Many (uttered) subsentences (clause fragments whose missing constituents are recoverable from context) clearly function like well-established clause types. It is prosody (or its typographic counterpart) that distinguishes interrogative ‘John?’ [‘Is that you?'] and imperative ‘JOHN!’ [‘Come in at once!'] from each other and from declaratives. An utterance of the name ‘John’ in answer to ‘Who’s that?’ must be judged declarative on prosodic and contextual grounds. The clipped clause ‘If only!’ is judged hypothetical on lexical and perhaps also contextual grounds. Such subsentences seem intuitively to function like distinct clause types and to have distinct PIs.

The primary illocution of a **declarative**, the default clause type, is defined in Table 1 below. In the definition the precondition (‘preparatory condition,’ Austin, 1975) establishes proper grounds for uttering a declarative clause. The illocutionary intention of a declarative arises directly from Speaker’s use of it,

given the precondition. As its primary illocution, the declarative performs an act of saying that Φ . The fact that declaratives (potentially) have truth values is the principal way to positively identify them. A declarative illocution is the primary illocution and not the illocutionary point of an utterance. There are few illocutionary points that cannot be achieved using a declarative: e.g., declaratives ‘I’m asking what your name is,’ ‘I’m telling you to stop making that noise’ issue directives. Performatives such as ‘I promise to be good,’ ‘You are prohibited from smoking’ also have the primary illocution of a declarative.

The PI of **interrogatives** is exemplified in Table 2. The purported option for the Hearer to refuse or plead inability to comply distinguishes interrogatives from imperatives – which purportedly offer no such option. These different options are what make interrogatives (askings) ‘more polite’ than imperatives (tellings) – but the pragmatics of the comparative uses of interrogatives and imperatives are too complex to be adequately discussed here.

Imperatives have the PI as shown in Table 3. Many imperative clauses are entreaties; they include supplications to someone powerful or wronged (‘Forgive me’), propositives (‘Let’s eat’), and instructions which are of primary benefit to Hearer (‘Have a nice day!’). Cross linguistically, first and third person ‘imperatives,’ often called ‘jussives,’ are frequently subjunctive (Palmer, 2001).

English has two independent **hypothetical** constructions that contrast with the three major clause types (see Table 4). Otherwise, in English, hypotheticals occur within the scope of declaratives, interrogatives, and imperatives, whose prosody they adopt along with their values. Hypotheticals may occur within the scope of a declarative, *T*, such that they

Table 1 The declarative PI

<form>	T [Φ], see (1). The typical prosody involves downdrift.
<description>	Speaker says that Φ . (Declaratives potentially have truth values.)
<precondition>	Speaker has reason to believe that Φ .
<illocutionary intention>	Speaker reflexively intends the utterance of the clause to be recognized as a reason for Hearer to believe that Φ . (Notice Speaker’s purported commitment to the truth of what is being said.)

Table 2 The interrogative PI

<form>	Q [Φ]. Lexical markers such as post clausal English ‘eh?’, clause final Japanese ‘ka’, English ‘wh’- words, ‘if’ and ‘whether’ in subordinate clauses, and/or morphosyntactic markers such as subject-auxiliary inversion in main clauses. Because rising intonation typically marks orientation toward Hearer, interrogatives typically display updrift.
<description>	Speaker asks Hearer something—makes a request (cf. Allan, 1986; questions are requests for information to be delivered verbally). There is a purported option for the Hearer to refuse (or plead inability) to comply.
<precondition>	Speaker purportedly believes that Hearer can or might be able to respond appropriately to what is asked in the clause.
<illocutionary intention>	Speaker reflexively intends the utterance of the clause to be taken as asking Hearer something.

(potentially) have truth values within some hypothetical world; for example, the complement of a verb of wishing. In T[I wish H[I were rich]], the 'were' is backshifted, i.e., looks like a past tense, but is in fact the nonpast derived from Old English subjunctive 'wære' (past tense would be 'I wish(ed) I **had been** rich'). In conditional I[H[If Harry should call]], tell him 'No', the protasis invokes a hypothetical world; the apodosis is imperative and supplies the PI. Within interrogatives, Q[H[Φ]], ('Could you mail this for me?'), the hypothetical is marked by one of four backshifted modals: 'could,' 'might,' 'should,' 'would.' What is here called *hypothetical* is often referred to as *subjunctive mood*; but the fact that it occurs within the scope of other moods makes this classification problematic—at least for those who wish to retain traditional definitions of mood.

Expressives are not traditionally recognized as a clause type because none are full syntactic clauses, e.g., 'Goodness gracious!' 'Wow!' 'Jesus!' 'Thanks.' 'Congratulations.' 'Hi.' 'Bye.' 'Good luck!' They can be distinguished on semantic and lexical-cum-morphosyntactic grounds; they are short and usually verbless. Like clauses, they can stand alone in expressing a complete thought. They are **idioms**, often with idiomatic counterparts in other languages. They do not admit passivization or negation. Expressives do not have truth values, so they are not declaratives; they do not make requests, so they are not interrogatives; they do not direct or entreat, so they are not imperatives. Some are subjunctive. Their principal function is to display a sometimes perfunctory, sometimes strongly felt emotive reaction or social interaction toward something that has occurred. One subset of them will serve to illustrate all.

'Thanks' is presumably foreclipped from 'I offer you my thanks.' However, one can't willy-nilly strip out the possessed noun from a string 'I offer (you) my NOUN' to create a synonymous utterance. The monosyllable 'Book' could, in a well-defined context such as proffering the book, be interpreted as 'I offer you my book'; but this not the kind of condition applicable to 'Thanks.' Furthermore one cannot foreclip 'I offer you my help' to create the monosyllable 'Help', which is imperative. Apparently, utterances with an expressive illocutionary point facilitate the creation of brief idiomatic counterparts for ritualized behavior and to satisfy the constraints on verbal ejaculation. 'Thank you' is normally subjectless and idiomatic by contrast with the declarative 'I/We thank you.' It is not part of a normal paradigm: It is not imperative like 'Thank him.' It cannot be tensed: 'Thanked you' is perhaps possible as an aphetic diary entry, but not as a normal report of past behavior. Both 'Thanks' and 'Thank you' are reported like any declarative ('I thanked him') – which is normal for expressives. 'No thanks' and 'No thank you' are idiomatizations of two separate subsentences ('No, thanks' / 'thank you') from which the disjuncture has disappeared. 'Thanks' and 'Thank you' mean 'I thank you'; their negation is 'I don't thank you.' This is not the meaning of 'No thanks' / 'thank you' which are explicit rejections of an offer; just as 'Yes, thanks' is an acceptance.

Exclamatives occur within the other clause types and do not count as a distinct clause type in their own right (despite the way they are discussed here). Exclamatives are all prosodically marked by exaggerated acoustic intensity (wide-ranging peaks and troughs) that is indicated orthographically by an exclamation mark, '!', and there is lexical marking of some

Table 3 The imperative PI

<form>	I[F[Φ]]. Prosodic marking: falling intonation. Imperatives occur only in independent clauses (including coordinate and appositive clauses). The verb is in the base form and non-past tense, but semantically future (F).
<description>	Speaker directs or entreats Hearer to do A. The direction or entreaty is present time, the act A is future – though often expected to start immediately. There is purportedly no option for the Hearer to refuse (or plead inability) to comply.
<precondition>	Speaker purportedly believes that Hearer can do A.
<illocutionary intention>	Speaker reflexively intends Hearer to take the utterance of the clause as a reason to do A.

Table 4 The hypothetical PI

<form>	H[Φ] where Φ is either 'Would that S' (e.g., 'Would that I were rich') or 'May S' ('May he rot in hell!') in which 'Would' and 'May' are main verbs using the root senses of these modals, and in which neither 'Would' nor 'May' can be negated.
<description>	Speaker vehemently wishes (hence the root senses) that the proposition in S may (or may have) come about in some hypothetical world accessible from the real world in which the utterance takes place.
<precondition>	Speaker has some reason to wish that S may (or may have) come about.
<illocutionary intention>	Speaker reflexively intends the clause to be taken as a reason for Hearer to believe that Speaker vehemently wishes that S may (or may have) come about in some hypothetical world.

declarative exclamatives by clause initial ‘how’ and ‘what’ immediately followed by the preposed clause constituent in the focus of exclamation (‘What a bloody whiner he was!’ ‘How she hated it!’). Speaker expresses within Φ a forcible, emphatic, or vehement point of approbation, delight, pleasure, pain, surprise, anger, irritation, etc.; the reflexive-intention is to focus Hearer’s attention on a part of what is being said in Φ as being of particular importance in the context of the communication in progress; but this is just a modification on the PI of the governing clause type. The same might be said of hypotheticals when they do not contrast with the other clause types, which is one reason to deny that there is a subjunctive mood in English.

Primary illocution (PI) is the semantics of clause type or corresponding subsentence. Declarative, interrogative, imperative, subjunctive, expressive, exclamative clause types or subsentences are identified through morphosyntactic, lexical, and prosodic form, their use in main and subordinate clauses, and behavior under negation. The PI is identified with Speaker’s illocutionary intention in using the clause type/subsentence. This PI is the starting point for Hearer inferring the illocutionary point of (Speaker’s message in) the utterance containing the clause, in the manner first described in Searle (1975) (see also Bach and Harnish, 1979; Allan, 1986, 2001). It is probable that all languages have most of the clause types/subsentences described here. Further research is needed to systematically accommodate conditionals, imperatives, jussives, laudatives, obligatives, optatives, prohibitives, propositives, and purposives within the network of relations manifest by clause types and illocutionary forces. Exactly how all these are to be related to the traditional category of mood remains to be seen.

See also: Compositionality: Semantic Aspects; Counterfactuals; Semantics of Interrogatives; Speech Acts; Type versus Token.

Bibliography

- Allan K (1986). *Linguistic meaning* (vols 1 and 2). London: Routledge and Kegan Paul (Repr. edn.: World Publishing Corporation, Beijing, 1991).
- Allan K (2001). *Natural language semantics*. Oxford, England & Malden, MA: Blackwell.
- Austin J L (1975). *How to do things with words* (2nd edn.). Urmson J O & Sbisà M (eds.). Oxford: Oxford University Press. [First published 1962].
- Bach K & Harnish R M (1979). *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.
- Bybee J L & Fleischman S (eds.) (1995). *Modality in grammar and discourse*. Amsterdam/Philadelphia: John Benjamins.
- Chafe W L (1995). ‘The realis-irrealis distinction in Caddo, the Northern Iroquoian languages, and English.’ In Bybee J L & Fleischman S (eds.). 349–365.
- Diogenes Laertius (1925). *Lives of eminent philosophers* (vol. 2). Hicks R D (trans.) Loeb Classical Library. London: Heinemann.
- Givón T (1994). Irrealis and the subjunctive. *Studies in Language* 18, 265–337.
- Householder F W (1981). *The syntax of Apollonius Dyscolus*. Amsterdam: Benjamins.
- Huddleston R & Pullum G K (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Jespersen O (1958). *The philosophy of grammar*. London: Allen & Unwin [First published 1924].
- Lancelot C (1644). *Novelle Méthode pour Comprendre, facilement et en peu de temps, la Langue Latine*. Paris.
- Lane A (1700). *A key to the art of letters*. London: facsimile. Menston: Scholar Press, 1969.
- Lunn P V (1995). ‘The evaluative function of the Spanish subjunctive.’ In Bybee J L & Fleischman S (eds.) 419–449.
- Mithun M (1995). ‘On the relativity of irreality.’ In Bybee J L & Fleischman S (eds.) 367–388.
- Mithun M (1999). *The languages of native North America*. Cambridge: Cambridge University Press.
- Palmer F R (2001). *Mood and modality* (2nd edn.). Cambridge: Cambridge University Press.
- Palmer F R (2003). ‘Modality in English: Theoretical, Descriptive and Typological issues.’ In Facchinetti R, Krug M & Palmer F R (eds.) *Modality in Contemporary English*. Berlin/New York: Mouton de Gruyter. 1–17.
- Quirk R, Greenbaum S, Leech G *et al.* (1985). *A comprehensive grammar of the English language*. London: Longman.
- Roberts J R (1990). ‘Modality in Amele and other Papuan languages.’ *Journal of Linguistics* 26, 363–401.
- Sadock J M (1974). *Toward a linguistic theory of speech acts*. New York: Academic Press.
- Sadock J M & Zwicky A (1985). ‘Speech Act Distinctions in Syntax.’ In Shopen T (ed.) *Language typology and syntactic description, vol. I: Clause structure*. Cambridge: Cambridge University Press. 155–196.
- Sanctius F (1585). *Minerva, seu de Causis Linguae Latinae*. Madrid.
- Searle J R (1975). ‘Indirect speech acts.’ In Cole P & Morgan J L (eds.) *Syntax and Semantics 3: Speech Acts*. New York: Academic Press. 59–82 [Repr. in Searle Expression and meaning: studies in the theory of speech acts. Cambridge: Cambridge University Press].
- Van Valin R D Jr & LaPolla R (1997). *Syntax: structure, meaning, and function*. Cambridge: Cambridge University Press.
- Wehmeier K F (2004). ‘In the mood.’ *Journal of Philosophical Logic* 33, 607–630.
- Whitney W D (1888). *A compendious German grammar* (6th edn.). New York: Holt, Rinehart and Winston.
- Wierzbicka A (1988). *The semantics of grammar*. Amsterdam: John Benjamins.

N

Natural Kind Terms

S Gross, University of Pennsylvania, Philadelphia, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Natural kind terms denote, or purport to denote, natural kinds – for example, physical, chemical, and biological kinds. They are of particular interest to philosophers concerned with what enables us to think and talk about natural kinds. According to *descriptivism*, the denotation of natural kind terms is determined by an associated description. According to *causal theories*, the denotation is determined by language users' causal interactions with exemplars. Hybrid views are also espoused. These positions parallel those taken on the denotation of proper names (cf. Stanley, 1997; see **Proper Names: Philosophical Aspects**).

Debates about natural kind terms carry implications for a variety of philosophical issues. Causal theories, for example, are *prima facie* in tension with a methodological individualism that assumes that cognitive properties – including the semantic properties of an individual's terms – are determined by properties intrinsic to the individual (cf. Burge, 1979). Discussions of natural kind terms also intersect with debates in cognitive psychology concerning tacit essentialism (cf. Gelman, 2003) and with discussions in the philosophy of science concerning the bases for successful inductive generalization (cf. Kornblith, 1993).

What Are Natural Kind Terms?

Human first languages do not syntactically or morphologically mark natural kind terms as such. Natural kind terms are distinguished solely by the kind of kinds they purport to denote. What then are natural kinds? The core idea is that they are object groupings that reflect real distinctions in nature – in particular, distinctions that play an explanatory role in (accurate) scientific theories. Natural kinds are contrasted with arbitrary groupings of objects whose behavior cannot be explained by shared fundamental

properties – for example, the kind comprising Plato's left ear, the natural numbers, and the shoes owned by current heads-of-state. Natural kinds are also contrasted with conventional groupings that reflect practical human concerns, such as the citizens of Pennsylvania. Relatedly, they are contrasted with artifactual kinds – objects grouped together in virtue of our intending them to fulfill certain functions, such as tables. Natural kind terms thus arguably include 'gold,' 'tiger,' and 'H₂O'; but not 'keys' or 'gorks' (where it is hereby stipulated that something is a gork if and only if it is Plato's left ear, a natural number, or a shoe owned by a current head-of-state).

This basic characterization leaves much room for disagreement about just which terms are natural kind terms. First, it admits adjectives – at least those that denote groupings that reflect real distinctions in nature (perhaps 'red' but not 'cultural' – cf. Kripke, 1972). But on some views, not all groupings are kinds, and so not all terms for them are kind terms. Many restrict kind terms to nouns or noun phrases that admit a generic reading, accept kind-level predicates like 'are widespread,' and are associated with criteria of individuation. These terms may be either mass or count nouns (cf. Carlson, 1991).

Second, there are disagreements about various further theses concerning natural kinds and, as a result, disagreements about which kind terms are natural kind terms. Some maintain, for example, that instances of a natural kind share an unchanging essence, that natural kinds fall into a non-cross-classifying hierarchy, or that natural kinds have no borderline instances. On such bases, they argue against the inclusion of various candidate kinds – and thus against the inclusion of various candidate kind *terms*. Debates concerning biological species and terms for them are a prominent example (cf. Dupré, 1981; LaPorte, 2004). (Philosophers debate many other issues concerning natural kinds, not all of which affect the classification of natural kind *terms*. For example, they disagree about whether natural kinds have reality independent of our categorizations and about what sorts of entities they are: sets, properties, complex individuals, or something *sui generis*.)

Third, the use of many putative kind terms seems sensitive to human interests. It may seem that we have discovered that the denotation of ‘water’ is shared by ‘H₂O.’ But the uses of ‘water’ and of ‘H₂O’ are sensitive to amounts and kinds of impurities in differing ways. We would hesitate to call tea ‘water’ but not the stuff flowing in a river, even though the latter might have more impurities vis-à-vis H₂O than the former. However, neither tea nor a contaminated river is H₂O. Some conclude that ‘water’ is not a natural kind term, while others argue that such phenomena can be explained pragmatically (cf. Malt, 1994; Braisby *et al.*, 1996; Abbott, 1997).

The Semantics of Natural Kind Terms: Descriptivism

We may distinguish three questions concerning the semantics of natural kind terms:

1. the *property* question: what are the semantic properties of natural kind terms?
2. the *foundational* question: in virtue of what do natural kind terms have these semantic properties?
3. the *competence* question: what is required for a speaker to grasp the semantic properties of natural kind terms?

The past 30 years of debate have been shaped to a large extent by Kripke’s (1972) and Putnam’s (1975) attacks on *descriptivism*. Classic descriptivism maintains three theses corresponding to these three questions. According to *property descriptivism*, the semantic properties of a natural kind term are those of some definite description. According to *foundational descriptivism*, a natural kind term has a particular semantic property – its denotation – in virtue of its relation to such a description. According to *competence descriptivism*, a speaker understands a term in virtue of associating it, explicitly or tacitly, with such a description.

Forms of descriptivism can be further articulated in various ways. For example, John Locke (1979: Book III, Chapter VI) – a *locus classicus* of descriptivism – in effect required that the relevant descriptions be constructed from observational terms. This is because his descriptivism is in service to his developmental and epistemological empiricism, according to which mental states arise from and are justified by experience. Locke distinguished the ‘real essence’ and the ‘nominal essence’ of substances, such as gold. The real essence is what makes gold what it is. Locke was pessimistic that we could ever discover a substance’s real essence. The nominal essence is the list of observable properties that a speaker associates with the term for this substance. Locke held that the term ‘gold’ signifies

an “abstract idea” composed of ideas of these observable properties (shiny, yellow, malleable, etc.). A natural gloss is that the content of a natural kind term like ‘gold’ is given by and identical to an observational description such as ‘the stuff that’s shiny, yellow, malleable, etc.,’ and a speaker grasps that this is the term’s content in virtue of taking that description to define it.

Against Descriptivism

Descriptivism is a natural view since such features clearly play a role in categorization. But Kripke and Putnam raised a host of objections aimed at separating how people categorize things from what determines the denotation in fact possessed by their terms.

First, there are *modal* objections. Consider ‘gold’ and a candidate Lockean description. Arguably, the sentence ‘gold is gold’ expresses a necessary truth – as does, on at least one reading, ‘the shiny, yellow (etc.) stuff is the shiny, yellow (etc.) stuff.’ But ‘gold is the shiny, yellow (etc.) stuff’ does not. For it is metaphysically possible both that there be stuff that possesses those observable properties but differs from gold in some fundamental way and that there be gold lacking some one of those properties. Relatedly, ‘it’s necessary that gold is gold’ expresses a truth, but ‘it’s necessary that gold is shiny, yellow, etc.’ does not. (It is thus often maintained that natural kind terms are rigid designators [see **Rigid Designation**], although how best to extend the concept beyond proper names is not obvious and depends on what one takes a kind to be; cf. Soames, 2002.)

A second, *epistemic* objection is that the descriptions associated with natural kind terms may be false without affecting the natural kind term’s denotation. Suppose that I falsely believe that tigers are the fastest cat and that this is a fundamental, defining characteristic of tigers. It does not follow that, when I use the word ‘tiger,’ I in fact denote cheetahs. More generally, one might reject a view according to which, as scientists revise their core beliefs in the light of new research, the denotation of their terms changes, so that scientists adopting different theories become communicatively isolated behind incommensurable vocabulary.

Finally, the *absence* objection points out that speakers competent in the use of a natural kind term often do not associate any particular description with it at all – or at least none that the speakers themselves would consider especially relevant semantically or that would uniquely pick out the natural kind at issue. Putnam (1975) notes that what he happens to know of beeches basically coincides with what he happens to know of elms, so he has available no

descriptions that would account for the distinct denotations of ‘beech’ and ‘elm.’ Further, even if there are descriptions that uniquely pick out these kinds, it is far from obvious that they play any role in a speaker’s cognitive economy.

Causal Theories

Rejecting descriptivism allows one to recognize the possibility of misalignment between what one’s terms are about and how one conceives of them. But, if not one’s conception, what determines the denotation of natural kinds terms? Causal theories deriving from Kripke (1972) and Putnam (1975) endorse a two-part account of the foundational semantics of natural kind terms: the denotation of natural kind terms is fixed by appropriate causal interaction with instances of the kind and can be socially transmitted to others by patterns of linguistic deference. A natural kind term is introduced with an intention – or comes to be used with an intention – that it denote stuff relevantly similar in underlying structure to *stuff like that*. What fixes the denotation is thus something like a Lockean real essence. ‘Gold’ denotes, not the stuff that shares various observable properties, but rather the stuff that has some fundamental properties responsible for the observable ones, whatever those fundamental properties might be. People not privy to samples of such stuff intend to use the term as do those who are. The denotation of ‘beech’ in Putnam’s idiolect, for example, is fixed by the fundamental properties of the samples with which those to whom he defers have interacted. Scientific inquiry enables us to learn about these fundamental properties, but linguistic competence does not require – indeed, arguably helps make possible – such knowledge.

On such a view, what semantic properties a natural kind term has are in part determined by features external to speakers – by the (perhaps unknown) underlying nature of the ostended stuff and by features of one’s linguistic community. Kripke’s and Putnam’s foundational semantics for natural kind terms thus supports *linguistic semantic externalism* – the thesis that the semantic properties of one’s terms do not supervene on one’s intrinsic states. A particularly vivid illustration is provided by Putnam’s ‘Twin Earth’ thought experiment. Suppose that there is another planet just like ours except that, in place of water (H₂O), its lakes, oceans, swimming pools, etc., are filled with XYZ, a liquid superficially similar to water in appearance, taste, odor, etc., but with a different chemical microstructure. Since XYZ is not H₂O, it is not water but rather something that merely resembles water. Now, suppose Oscar is a denizen of Earth, and Twin Oscar is his counterpart on Twin Earth – an

exact duplicate with respect to intrinsic properties. Out of Oscar’s mouth, ‘water’ denotes water (H₂O), but not XYZ. Were it pointed out to Oscar that the liquid he just referred to as ‘water’ was in fact made up of completely different stuff, he would admit that he’d been misled by its superficial resemblance to water. Out of Twin Oscar’s mouth, ‘water’ denotes XYZ (that is, what *they* call ‘water’ on Twin Earth), but not water – for parallel reasons. But Oscar and Twin Oscar are intrinsically identical. Therefore, their words’ semantic properties are a function, not just of their intrinsic states, but also of external factors. Such externalism stands in sharp contrast to, for example, Chomsky’s internalist conception of language (see *E-Language versus I-Language*).

Descriptivism Redux

Descriptivists have defended themselves both by replying to Kripke’s and Putnam’s objections and by raising problems for causal theories.

In reply to the *modal* argument, one finds four objections, presented here in increasing order of concessiveness. Some reject the very idea of metaphysical modalities. Others reject the intuitions behind the particular modal claims made, sometimes adding that such intuitions are sensitive to framing effects (and might not be cross-culturally robust – cf. Machery *et al.*, 2004). Then there are those who deny that critics have considered the *right* descriptions: if ‘gold’ is associated with a *rigidified* description such as ‘what’s actually the most precious yellow metal,’ the modal argument is blocked. Finally, some point out that the argument, if successful, applies only against *property* descriptivism and so does not yet pose a problem for strictly *foundational* descriptivists.

In reply to the *epistemic* objection, a descriptivist might likewise complain that causal theorists focus on the wrong descriptions. There must be *some* accurate description that can be associated with the natural kind term. For instance, the relevant description could be constructed from the causal theorist’s account of the determination of natural kind terms’ denotation: ‘the stuff that (actually) stands in such-and-such causal relations to users of the term K.’ It may be objected that speakers can be competent in the use of natural kind terms despite not associating them with such theoretically sophisticated descriptions. We return to this sort of objection presently. But there are also less theoretically sophisticated candidate descriptions, such as ‘the stuff that’s (actually) relevantly like *this* stuff, or the stuff others have talked about’ (so long as there is no bar on demonstratives), or on some views ‘the stuff that’s called water in this language’ (cf. Haas-Spohn, 1997).

In reply to the *absence* objection, a descriptivist might posit tacit knowledge of the description's bearing on the natural kind term. That is, the descriptivist might hypothesize that the mind or brain associates natural kind terms with such descriptions, even though a speaker is not necessarily aware of this. Alternatively, a descriptivist might argue that such descriptions can play the role required by *property* and/or *foundational* descriptivism despite speakers' ignorance of them. Rather than positing tacit knowledge, the descriptivist might then abandon *competence* descriptivism altogether. This might appeal in particular to those who deny that a semantic theory must also play a role in an account of semantic competence. According to such theorists, the task of semantics is to characterize the semantic properties of expressions, not to explain how speakers grasp those properties (cf. Katz, 2004).

In addition to replying to causal theorists' objections, descriptivists emphasize that causal theories face *prima facie* problems of their own. First, it is not obvious how causal theories should accommodate apparently denotationless natural kind terms such as 'phlogiston' and 'unicorn' (see **Empty Names**). These terms are not synonymous. But it is unclear how a causal theorist can distinguish them semantically, since the theorist cannot appeal to a difference in denotation. Descriptivists face no problem here, since they can associate such terms with nonsynonymous descriptions ('the stuff flammable material gives off in combustion'; 'the horse-like mammal with a magical horn').

Moreover, causal theorists face the *prima facie* problem that, on their view, 'phlogiston' is not a denotationless term after all, since it was introduced via interaction with instances of combustion and thus arguably denotes *oxygen*, the stuff in fact responsible for the observed behavior. Unless they can resist this consequence, causal theorists would be committed to a history of science according to which phlogiston theorists were correct that phlogiston exists but simply wrong that it comes from the flammable material.

Finally, an object can be an instance of various natural kinds. If my use of 'gold' is associated with exemplars that are not only Au but also all instances of a particular isotope, what determines to which kind my term refers? Perhaps indeed the term should denote *metals*. This is an instance of the *qua*-problem (anticipated by Locke – cf. Stanford, 1998): it is unclear of what kinds the instances are exemplars, in particular whether we are to consider the instance *qua* gold, *qua* metal, *qua* a certain isotope, etc. Here too it can seem that the foundational semantics requires a constraining descriptive component.

Hybrid Views

In part for these reasons, some have developed hybrid views. There are two not mutually exclusive ways one can combine a moderate descriptivism with a causal theory – both of which have antecedents in Putnam (1975). First, some defend a limited *property* descriptivism while abandoning the claim that the denotation of natural kind terms is determined solely by such descriptions. The idea is that among the semantic properties of a natural kind term is an associated description (or a meaning with description-like content), even though the denotation is not determined by this description and might not align with its extension. Second, some maintain that a natural kind term's denotation is determined by causal relations to exemplars as constrained by some descriptively characterized requirement. It is not sufficient to say that a natural kind term denotes *stuff like that*: it must be stuff that is *relevantly similar* to the exemplar in ways that require at least some descriptive elaboration (cf. Stanford and Kitcher, 2000).

Neodescriptivists have attempted to absorb the insights of causal theories in other ways as well. We have already noted that one might include rigidifying and/or demonstrative expressions in the descriptions. In addition, some theorists distinguish two dimensions of semantic content: a description-like first intension and a second intension that, in the case of natural kind terms, is a function of the first intension and the actual environment of the speaker. The first intension determines a classically descriptivist extension. The second intension, however, does not determine an extension on its own at all; it constrains an extension that is in part determined by how the world in fact is (cf. Chalmers, 2002; see **Two-Dimensional Semantics**).

See also: Causal Theories of Reference and Meaning; Descriptions, Definite and Indefinite: Philosophical Aspects; E-Language versus I-Language; Empty Names; Extensionality and Intensionality; Mass Nouns, Count Nouns, and Non-count Nouns: Philosophical Aspects; Meaning: Development; Nominalism; Proper Names: Philosophical Aspects; Reference: Philosophical Theories; Rigid Designation; Two-Dimensional Semantics.

Bibliography

- Abbott B (1997). 'A note on the nature of 'water.' *Mind* 106, 311–319.
- Atran S (1998). 'Folk biology and the anthropology of science: cognitive universals and cultural particulars.' *Behavioral and Brain Sciences* 21, 547–609.
- Braisby N, Franks B & Hampton J (1996). 'Essentialism, word use, and concepts.' *Cognition* 59, 247–274.

- Burge T (1979). 'Individualism and the mental.' *Midwest Studies in Philosophy* 4, 73–121.
- Carlson G (1991). 'Natural kinds and common nouns.' In von Stechow A & Wunderlich D (eds.) *Semantik/ Semantics*. Berlin: Walter de Gruyter. 370–398.
- Chalmers D (2002). 'On sense and intension.' *Philosophical Perspectives* 16, 135–182.
- Dupré J (1981). 'Natural kinds and biological taxa.' *Philosophical Review* 90, 66–90.
- Gelman S (2003). *The essential child: origins of essentialism in everyday thought*. Oxford: Oxford University Press.
- Haas-Spohn U (1997). 'The context-dependency of natural kind terms.' In Künne W, Anduschus M & Newen A (eds.) *Direct reference, indexicality and propositional attitudes*. Paolo Alto: CSLI Publications. 333–349.
- Katz J (2004). *Sense, reference, and philosophy*. Oxford: Oxford University Press.
- Kornblith H (1993). *Inductive inference and its natural ground*. Cambridge: MIT Press.
- Kripke S (1972). 'Naming and necessity.' In Davidson D & Harman G (eds.) *Semantics of natural language*. New York: Humanities Press. 253–355. [Reprinted with additional material as *Naming and necessity*. Cambridge: Harvard University Press, 1980.]
- LaPorte J (2004). *Natural kinds and conceptual change*. Cambridge: Cambridge University Press.
- Locke J (1689). In Nidditch P (ed.) *An essay concerning human understanding*. Oxford: Oxford University Press [first published in 1689].
- Machery E, Mallon R, Nichols S & Stich S (2004). 'Semantics, cross-cultural style.' *Cognition* 92, B1–B12.
- Malt B (1994). 'Water is not H₂O.' *Cognitive Psychology* 27, 41–70.
- Putnam H (1975). 'The meaning of "meaning."' In Gunderson K (ed.) *Language, mind, and knowledge*. Minneapolis: University of Minnesota Press. 131–193.
- Soames S (2002). *Beyond rigidity*. Oxford: Oxford University Press.
- Stanford P K (1998). 'Reference and natural kind terms: the real essence of Locke's view.' *Pacific Philosophical Quarterly* 79, 78–97.
- Stanford P K & Kitcher P (2000). 'Refining the causal theory of reference for natural kind terms.' *Philosophical Studies* 97, 99–129.
- Stanley J (1997). 'Names and rigid designation.' In Hale B & Wright C (eds.) *A companion to the philosophy of language*. Oxford: Blackwell. 555–585.

Natural versus Nonnatural Meaning

A Barber, The Open University, Milton Keynes, UK

© 2006 Elsevier Ltd. All rights reserved.

Grice's Distinction

H. P. Grice noticed a potentially confusing lexical ambiguity in the word 'means' (see Grice, 1957). Contrast claim (1) with claim (2):

- (1) The flooded river means that the mountain snow is melting
- (2) The sentence 'La neige des montagnes fond' means that the mountain snow is melting

If it turned out that the mountain snow was not melting, then claim (1) would have to be withdrawn, whereas claim (2) would not. The French sentence in (2) would mean what it does even if it was untrue. Grice distinguishes what he calls natural meaning, which is what 'means' means in the context of a claim such as (1), from nonnatural meaning, invoked in (2). He offers a range of criteria for distinguishing the two kinds of meaning – 'meaning_N' and 'meaning_{NN}' as he abbreviates them – but the key difference is sometimes put in terms of meaning_N, unlike

meaning_{NN}, being *factive*: what is meant_N can only be a fact, whereas what is meant_{NN} need not.

Sentences are not the only kind of entity that may be said to have meaning_{NN}. Others include people, as when we say that so-and-so meant something; utterances of a sentence as opposed to the sentence itself; and mental states, such as beliefs, which have representational content without necessarily representing accurately. The distinction seems to be available no matter what one takes to be the primary bearers of meaning_{NN}.

Grice stressed this distinction because he was interested in nonnatural meaning, which he deemed to be harder to pin down and so more in need of investigation than natural meaning. Natural meaning is simply a relation of necessitation between one event and another, for example a causal relation such as the one alleged in (1). The investigation of nonnatural meaning is handicapped, he felt, by confusions that arise through failure to see the distinction between it and meaning_N. In his 1957 paper, he criticizes the attempt to understand meaning_{NN} in terms of meaning_N, and then appeals to the distinction in the course of blocking objections to his own theory of meaning_{NN}.

Grice's Theory of Non-natural Meaning

Grice's distinction is acknowledged even by those who do not endorse his theory of meaning_{NN}. But this latter has been influential nonetheless (see Schiffer, 1972 for a development of it, and Schiffer, 1987 for a retraction). The meaning_{NN} of an act, he suggests, turns entirely on the intention of its agent. Not all acts have meaning_{NN}, but those that do are performed with a triplex of intentions. First, the agent intends to bring about an effect of some kind in the mind of her audience – a change of belief, for example. Second, she intends to be recognized by this audience as having this first intention. Third, she intends the effect to arise *through* this recognition. If an act is performed with intentions of this form, then the intended effect is said by Grice to be the act's meaning_{NN} (or, equivalently, it is what the agent meant_{NN} by the act; or again, it is what the expression used means_{NN} on that occasion).

In practice, recognition of the intention is usually achievable only because of the close association between acts typed according to whether they share the same meaning_{NN} and acts typed according to an overt, recognizable pattern. In the linguistic case, for example, typing according to the sentence uttered will usually group together utterances with the same meaning_{NN}. Shared knowledge of which sentences are paired with which meanings_{NN} is what makes individual acts of meaning_{NN} possible in the first place. This leads Grice to introduce a derivative notion, timeless meaning_{NN}, which is the meaning associated with the repeated use of a particular pattern. Sometimes this timeless meaning_{NN} of a sentence is distinct from the meaning_{NN} of an utterance of it on an occasion. Someone could use the sentence *I know where you live* to mean_{NN} something distinct from what it timelessly means_{NN}. It is no coincidence that Grice was among the first to notice and develop a theory of the phenomenon of implicature (see Grice, 1975).

Other Remarks

Though the existence of a distinction between natural and nonnatural meaning is clear once pointed out, it is easy to become confused over its precise nature. After all, utterances are events and, as such, have meaning_N as well as meaning_{NN}. And (3) is true on both readings of 'means':

- (3) The canyon-dweller's shout of "Here comes an echo" meant that we would shortly hear an echo

Moreover, the utterances of an omniscient truth-teller – God, for example – mean_N whatever they mean_{NN}. Efforts by philosophers of mind and language in the 1980s to understand systems of representation naturalistically (see Millikan, 1984; Dretske, 1988) may shed light on these cases. A distinction is sometimes drawn between what an instance of mental representation *R* *indicates* and what it *represents*. What it indicates (or what information it carries) is understood entirely in terms of the causal relations instances of *R* tend to enter into – something like meaning_N. It represents not what it indicates but *what it is the function* of instances of *R* to indicate, within the representational system as a whole. So when the representation system, i.e. the mind or some part of it, is functioning properly, an instance of *R* will indicate what it represents. Misrepresentation is a matter of functional breakdown in the system, resulting in instances of *R* that fail to indicate what they represent, i.e., what it is their function to indicate. An omniscient truth-teller can be thought of as a properly functioning representational system, which is why God's utterances mean_N (or indicate) what they mean_{NN} (or represent).

See also: Expression Meaning versus Utterance/Speaker Meaning; Implicature; Intention and Semantics; Representation in Language and Mind.

Bibliography

- Dretske F (1988). *Explaining behavior: reasons in a world of causes*. Cambridge: MIT Press.
- Grice H P (1957). 'Meaning.' *Philosophical Review* 66, 377–388.
- Grice H P (1975). 'Logic and conversation.' In Cole P & Morgan J (eds.) *Syntax and semantics*, vol. 3. London: Academic Press. Reprinted in H. P. Grice, *Studies in the way of words*. Cambridge: Harvard University Press, 1989.
- Millikan R G (1984). *Language, thought, and other biological categories: new foundations for realism*. Cambridge: MIT Press.
- Schiffer S R (1972). *Meaning*. Oxford: Clarendon Press.
- Schiffer S (1987). *Remnants of meaning*. Cambridge: MIT Press.

Naturalism

J Kulvicki, Dartmouth College, Hanover, NH, USA

© 2006 Elsevier Ltd. All rights reserved.

Philosophical naturalism is the view that traditional problems in epistemology, language, metaphysics, mind, and morality can be solved only by appeal to the methods and results of empirical research. We cannot hope to solve these problems *a priori*, by reflecting on our everyday concepts and the logical relations between them. On the face of it, naturalism seems quite appealing, and almost obvious. No one thinks we can figure out how the sun works without investigating the matter empirically. Why would we think philosophy's Big Problems will not yield to science's prying eyes?

Naturalism is controversial because many of these topics are, broadly speaking, normative. The empirical sciences describe, but they do not prescribe: they have little to say about how the world *should* be. As a result, the sciences leave out precisely the normative aspects that make these phenomena so interesting, valuable, and puzzling. Naturalists take up this challenge in radical and moderate forms. Moderates try to fit norms into the natural world by explicating the conditions for satisfying a norm in nonnormative, scientifically informed terms. Radicals refuse norms altogether. Normative inquiry is a lost cause, say the radicals, so it should be abandoned in favor of purely descriptive endeavors.

Epistemology asks what knowledge is and how we should go about getting it. What makes certain methods for learning the truth reliable? Under what conditions are we justified in believing something? Naturalized epistemology treats these as scientific questions. We need to know how our perceptual systems work, how we reason, and how we make decisions in order to evaluate competing approaches to learning. The psychological and brain sciences thus become relevant to the theory of knowledge. In the last century Willard v. O. Quine, a radical, is considered the father of naturalized epistemology (1969). Moderates, such as David Armstrong (1961), Alvin Goldman (1967), and Fred Dretske (1968, 1981), were influential in making science part of the solution to the problems of knowledge. Kornblith (1994) provided a nice collection of essays on naturalizing epistemology, and Kitcher (1992) provided a valuable discussion of naturalism's history and current place in epistemology.

By contrast, H. H. Price claimed that “[e]mpirical science can never be more trustworthy than perception, upon which it is based. . .” (1932: 1). Price and the other sense datum theorists of the first half of the

20th century argued that since knowing the results of science depends on perceiving the world, we cannot use the science of perception as a guide to epistemology. They were naturalists in that they sought to explain knowledge in terms of what we sense and what we cannot in fact doubt, the latter being non-normative notions (cf. Kim, 1993). They were not naturalists, in that they thought we must first secure foundations for perceptual knowledge before we can trust science as a tool for furthering or improving it. Epistemology thus becomes an *a priori* endeavor. The sense datum theorists in particular sought that which we cannot doubt – for Bertrand Russell it was “facts of sense (i.e., of *our own* sense-data) and the laws of logic” (1914: 119) – and tried to show how to build further knowledge thereupon by logical construction. This program culminated in Rudolf Carnap's *Der Logische Aufbau der Welt* (1928), an impressive but failed effort, at least given general opinion.

Quine insisted that “it is within science itself, and not in some prior philosophy, that reality is to be identified and described” (1981: 21). A radical naturalist, Quine saw epistemology as a purely descriptive, nonnormative enterprise that asks how we in fact go about acquiring beliefs and knowledge about the world (Kim, 1993).

If we are out simply to understand the link between observation and science, we are well advised to use any available information, including that provided by the very science whose link with observation we are trying to understand. (Quine, 1969: 76)

Quine's influential ‘Two dogmas of empiricism’ (1951) provided the groundwork for this view. First he denied that all our knowledge derives from a privileged class of claims about, for example, what we sense. The sense datum theorists thought that claims about what we sense are individually indubitable. From these special claims, we derive the rest of our knowledge. Quine, by contrast, thought that there is no special class of individually confirmed claims. Claims are not confirmed one by one but *en bloc*. “Our statements about the external world face the tribunal of sense experience not individually but only as a corporate body” (Quine, 1951: 43). If this is so, we would do well to bring the results of science into the discussion of what knowledge is and how we actually go about getting it. Quine believed, after all, that the senses provide us with all the information we can get about the world, but he saw this as a result of scientific research, not a claim prior to science. Second, Quine (1951, 1954) rejected the idea that any claims are analytic, or true just in virtue of their meaning and not in virtue of whether they fit with

observation (see also Burge, 1992: 6–9). Statements that seem to be true just in virtue of their meaning, and even Russell's 'laws of logic,' are not in principle immune to disconfirmation any more than ordinary empirical claims, so there is little reason to exclude the latter from one's account of knowledge.

Quine was criticized for straying from his naturalistic scruples when it came to language. Based partly on worries about language acquisition, Quine argued in favor of two theses: the indeterminacy of translation (1960) and the inscrutability of reference (1960, 1970). According to the former, there is no fact of the matter about which translation among competing translations of a given language into another is correct. The claim is not that it is rather difficult to be sure one has achieved the proper translation but that there is no fact of the matter about which is correct at all. The latter thesis claims that there is no fact of the matter about what specific terms in a language refer to. Taken together, these theses led Quine to a general skepticism about meaning and reference.

Chomsky (1968, 1975, 1980) objected to Quine's theses on the simple grounds that they contradicted Quine's claims about every other science. Just about everyone has agreed that the empirical evidence underdetermines one's theory of a phenomenon. So, indefinitely many theories can account for any given set of evidence. This does not interfere with science's progress, and it is not clear why the study of language should answer to a higher standard than the other sciences.

If the underdetermination of physical theory by evidence does not lead us to abandon the 'realistic point of view' [...] then the comparable underdetermination of grammatical theory by evidence does not support Quine's claim that there is no fact of the matter [...] to be right or wrong about . . . (Chomsky, 1975: 183)

Complaints in a similar vein can be found in Fodor and Lepore (1992: 70–82). Quine (1960: 75–6, 1968) argued that the cases are not parallel. In a remark that nicely paralleled Quine's quote (1969: 76), Robert Fogelin complained:

... Quine's attempt to trace a path from stimulation to science is carried out with almost no concern for the psychological reality of the process, that is, with almost no references to empirical investigations of how language is actually acquired. (1997: 561)

In the end, Quine's controversial claims should, by his own lights, be subject to empirical confirmation or disconfirmation. According to linguists and many philosophers, the weight of the evidence seems to be against him. Scott Soames (2003: 404–405)

suggested that Quine's views might not admit of empirical support one way or the other, which would, on Quinean grounds, render them meaningless and useless.

Naturalized philosophy of mind tries to show how facts about intentionality, consciousness, meaning, and the like can be explained in terms of what the natural sciences – particularly psychology, neuroscience, and biology – tell us about ourselves. Here, too, we find problems with normativity. For example, concepts have intentional contents. They are about things and states of affairs. What, then, makes my concept of snow about snow? Perceptual encounters with snow cause me to think of snow and have certain beliefs about it. Perhaps, then, my concept of snow has its content in virtue of what causes me to deploy it. Cleverly placed whipped cream, however, can fool me into believing that there is snow on my windowpane. Whipped cream is not snow, of course, and my snow concept is not about whipped cream at all. It cannot be just anything causing me to deploy my snow concept figures in its content. If it did, the content would be quite disjunctive – snow or whipped cream or cotton gauze at a distance, etc. – and none of my uses of the concept would be mistaken. Any naturalistic theory of content must explain how we incorrectly use concepts, and this appeals to norms regulating their application. Scientific theories tend not to appeal to norms. Attempts to naturalize content are due to Fred Dretske (1981), Ruth Millikan (1984), and Jerry Fodor (1990), among others. Burge (1992) helpfully reviewed the development of naturalized philosophy of mind and language in the 20th century (see **Representation in Language and Mind**).

Chomsky's (1980, 1995, 2000) commitment to naturalism has been 'methodological' in that he has taken scientific explanation to be the goal when we investigate language and mind. He famously denied that there is an interesting 'metaphysical' sense to naturalism according to which research aims to explain mental and linguistic phenomena in physical or biological terms. Chomsky has been skeptical of metaphysical naturalism because, first, reduction is rarely a prerequisite for conducting science. For example, biology has not been reduced to physics, but it is a science nonetheless. And second, we do not have a clear enough conception of the physical for the metaphysical project to make sense.

Lacking a concept of 'matter' or 'body' or 'the physical' we have no coherent way to formulate issues related to the 'mind-body problem' ... the sciences postulate whatever finds a place in intelligible explanatory theory, however offensive that may be to common sense. (Chomsky, 1995: 4–5)

Naturalists are committed to following science wherever it leads, but we do not understand the world well enough to make a stronger, metaphysical commitment to reducing the mind to the physical. Since our notion of what is physical is frequently in flux, it seems odd to insist that the mind must be explicable in terms of it. Responses to Chomsky's views, with replies by Chomsky, can be found in Antony and Hornstein (2003).

Biology, more than most other sciences, must deal with normative notions. What is the function of this trait? Is it adaptive? What was its original purpose? And so on. Philosophers have not missed the fact that biology offers a promising approach to the moderate naturalists' goals. Alex Rosenberg (1996: 4) suggested that "[i]f naturalism is to replace *a priori* first philosophy with scientific theory, then at least for the present the theory in question will be Darwin's." Millikan's work (1984, 1993) has focused on explicating a naturalistically acceptable notion of biological function and applying it to problems in the philosophy of mind and language. Dretske (1988), David Papineau (1993), Karen Neander (1995), and Daniel Dennett (1995), among others, have followed her lead.

See also: Representation in Language and Mind.

Bibliography

- Antony L & Hornstein N (eds.) (2003). *Chomsky and his critics*. Malden, MA: Blackwell.
- Armstrong D (1961). *Perception and the physical world*. London: Routledge.
- Burge T (1992). 'Philosophy of Language and Mind: 1950–1990.' *Philosophical Review* 101(1), 3–51.
- Carnap R (1928). *Der Logische Aufbau der Welt*. Berlin: Felix Meiner.
- Chomsky N (1968). 'Quine's empirical assumptions.' *Synthese* 19, 53–68.
- Chomsky N (1975). *Reflections on language*. New York: Pantheon Books.
- Chomsky N (1980). *Rules and representations*. Oxford: Blackwell.
- Chomsky N (1995). 'Language and nature.' *Mind* 104(413), 1–61.
- Chomsky N (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Dennett D (1995). *Darwin's dangerous idea*. New York: Harper & Row.
- Dretske F (1968). *Seeing and knowing*. London: Routledge.
- Dretske F (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Dretske F (1988). *Explaining behavior*. Cambridge, MA: MIT Press.
- Fodor J (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.
- Fodor J & Lepore E (1992). *Holism: a shopper's guide*. Cambridge, MA: MIT Press.
- Fogelin R (1997). 'Quine's limited naturalism.' *Journal of Philosophy* 94(11), 543–563.
- Goldman A (1967). 'A causal theory of knowing.' *Journal of Philosophy* 64, 357–372.
- Kim J (1993). 'What is "naturalized epistemology"?' In *Supervenience and mind*. Cambridge: Cambridge University Press.
- Kitcher P (1992). 'The naturalists return.' *Philosophical Review* 101(1), 53–114.
- Kornblith H (ed.) (1994). *Naturalizing epistemology*. Cambridge, MA: MIT Press.
- Millikan R (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
- Millikan R (1993). *White Queen psychology and other essays for Alice*. Cambridge, MA: MIT Press.
- Neander K (1995). 'Misrepresenting and malfunctioning.' *Philosophical Studies* 79, 109–141.
- Papineau D (1993). *Philosophical naturalism*. Oxford: Blackwell.
- Price H H (1932). *Perception*. London: Methuen.
- Quine W (1951). 'Two dogmas of empiricism.' In Quine (1961). 20–46.
- Quine W (1954). 'Carnap on logical truth.' In Quine (1976). 107–132.
- Quine W (1960). *Word and object*. Cambridge, MA: MIT Press.
- Quine W (1961). *From a logical point of view* (2nd edn.). Cambridge, MA: Harvard University Press.
- Quine W (1968). 'Reply to Chomsky.' *Synthese* 19, 274–283.
- Quine W (1969). 'Epistemology naturalized.' In *Ontological relativity and other essays*. New York: Columbia University Press.
- Quine W (1970). 'On the reasons for indeterminacy of translation.' *Journal of Philosophy* 67, 178–183.
- Quine W (1976). *The ways of paradox and other essays*. Cambridge, MA: Harvard University Press.
- Quine W (1981). 'Things and their place in theories.' In *Theories and things*. Cambridge, MA: Belknap. 1–23.
- Rosenberg A (1996). 'A field guide to recent species of naturalism.' *British Journal for the Philosophy of Science* 47(1), 1–29.
- Russell B (1914). *Our knowledge of the external world*. Open Court.
- Soames S (2003). *Philosophical analysis in the twentieth century*, vol. 1. Princeton: Princeton University Press.

Negation: Philosophical Aspects

D DeVidi, University of Waterloo, Waterloo, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Most people, if they think of negation at all, have in mind something fairly simple, namely classical (or Boolean) negation. Classical negation is an operator that works as a sort of toggle switch on propositions or predicates: When applied to a proposition with the semantic value 'True,' it returns a proposition with the semantic value 'False' and vice versa. When applied to a predicate true of some collection of individuals (of some type), it returns another predicate that is true of exactly those individuals (of that type) for which the original failed to hold. If one regards relations as predicates that apply to pairs (triples, quadruples, etc.), and considers pairs (etc.) as individuals of a higher type, one has a general-purpose notion of negation. Make the convenient assumption that every proposition has exactly one of the two truth values just mentioned, and that predicates have sharply defined domains of application and sharp boundaries within those domains, and nothing more needs to be said.

It is increasingly rare to find philosophers of logic who are happy with the idea that this is the whole story about negation. Most philosophical issues regarding negation are illuminated by considering some reasons for their dissatisfaction. However, first it is useful to consider the question of what properties an operator needs to have to count as a genuine negation. This is surely the fundamental philosophical question about negation, and it must be answered before one can adjudicate the justice of the most common complaint raised against various nonclassical negations – that they are not really negations at all.

To keep the discussion manageably brief, we gloss over some philosophical niceties. For instance, we shall move freely between negations of propositions and of predicates when it makes examples clearer, although some contend that these might not involve the same operator. We also mostly ignore the question of whether negation is an operation on sentences (and predicates) or on propositions (and properties), thereby pretending some issues are simpler than they really are. We use the symbol ' \neg ' for the negation operator.

Possible Properties of Negation

Probably the strongest opinion most people would have about negation, is that for any sentence P , P and $\neg P$ cannot both be true. If $\neg P$ is really the negation of P , then it must be incompatible with P . However, one

must remember the distinction between negations and mere contraries. For if P is ' x is white,' then ' x is black' is certainly incompatible with P – it is a contrary of P – but it is not $\neg P$. For, surely, if the x in question is orange, $\neg P$ must come out true. However, ' x is black' does not. One natural way to think about $\neg P$ is as the weakest proposition that is incompatible with P . In our example, $\neg P$ is a proposition implied by any proposition that rules out x being white, perhaps equivalent to ' x is black or x is orange or x is yellow ...,' running through all the nonwhite colors. Thinking of negation in this way, and assuming that if A is incompatible with B then also B is incompatible with A (which not everyone grants, for reasons we will not pursue), is equivalent to assuming that negation obeys the following law: A implies $\neg B$ if and only if B implies $\neg A$. (In this view, A implies $\neg B$ exactly when A is incompatible with B . If this means also that B is incompatible with A , this happens exactly when B implies $\neg A$.) This is the characteristic law of so-called 'minimal negation,' and almost all candidate negations satisfy it.

Of course, the classical negation of P is not merely the weakest proposition contrary to P (although it is that) but also satisfies many additional laws. Most significantly, $\neg P$ must be a complement of P ; that is, not only must $(P \text{ and } \neg P)$ be a logical falsehood but also $(P \text{ or } \neg P)$ must be a logical truth. That $\neg P$ satisfies the latter condition for any P is called the law of excluded middle and is close enough kin to the principle of bivalence (for any P , P gets exactly one of two truth values, and so $\neg P$ gets the other) that we can ignore the difference here. Other important principles that hold for classical negation include the four De Morgan laws: $\neg(P \text{ and } Q)$ implies $(\neg P \text{ or } \neg Q)$ and vice versa, and $\neg(P \text{ or } Q)$ implies $(\neg P \text{ and } \neg Q)$ and vice versa. The logical principle sometimes called 'explosion' is valid for classical negation: Everything follows from a contradiction (i.e., a pair $P, \neg P$). In the presence of explosion, the law of excluded middle is equivalent to the principle of double negation elimination: $\neg\neg P$ implies P , which therefore also holds for classical negation.

We can now efficiently describe the most discussed nonclassical negations. (The philosophical significance of these distinctions should become clear in later sections.) If a minimal negation also satisfies explosion, we have an 'intuitionistic negation.' On the other hand, if a minimal negation satisfies the four De Morgan laws, we have what is sometimes called 'Ockham negation.' This unfortunate name is needed because common usage has it that a 'De Morgan negation' is an Ockham negation for

which also the principle of double negation elimination holds. An 'orthonegation' is both intuitionistic and De Morgan. Classical negation is orthonegation in a logical system in which 'and' distributes over 'or' and vice versa. Nonclassical orthonegations show up in nondistributive quantum logics. Distributivity guarantees that every proposition has at most one complement, but we will ignore the differences between classical and orthonegation here.

One sometimes sees discussion of 'subminimal negation.' It is characterized by the condition that the following principle of contraposition must hold: If A implies B , then $\neg B$ implies $\neg A$. Every minimal negation satisfies this principle, but not conversely. Although most would accept that this is a necessary condition for an operator to count as a negation, it seems insufficient because there are operations in natural language that seem to be subminimal negations that are difficult to take seriously as negations. Skipping details, an example is the operation we might call 'few N ': If ' x is a fish' implies ' x has gills,' then 'Few dogs have gills' implies 'Few dogs are fish,' and similarly for any similar example, but 'few N ' looks even less like a negation than any mere contrary forming operator. Sometimes, especially in programming applications, one sees operations called 'negations' that are not even subminimal (e.g., some 'negation as failure' operators). We shall not consider these operations.

Negations, Consistency, and Paradoxes

We begin with some responses to the so-called 'semantic paradoxes,' especially the liar paradox. Consider (L):

(L): L is false.

(L) certainly looks to be true if and only if it is false. Nonclassical negations often appear in proposed solutions to such paradoxes; we briefly consider a simple example later. For now, we consider so-called 'dialetheism.' Dialetheists hold that some sentences P are both true and false (and so that both P and $\neg P$ are true, if $\neg P$ is true whenever P is false). Sentences such as (L) are their prime candidates for this status. In barest outline, they argue that a century of intense effort aimed at explaining what is going on with (L) while preserving consistency has led, at best, to solutions that are plausible only to those who formulate them. The mistake, they suggest, is in trying to preserve consistency.

One obvious objection to dialetheism is that if P and $\neg P$ can both be true at the same time, the \neg symbol is not functioning as a negation. In such a case, $\neg P$ is not even a contrary of P , violating the most basic intuition

about negation. One dialetheist response is that in many dialetheic logics, $\neg(P \text{ and } \neg P)$ is a necessary truth for all P , even those where P is both true and false. They suggest that this captures the truth behind the idea that negations are contraries. It is just a surprising development that $(P \text{ and } \neg P)$ is also true in some cases.

Dialetheist must defend a negation for which explosion is not valid, on pain of holding that all statements are true. Others defend nonexplosive logics, called 'paraconsistent' logics, for a variety of other reasons. For instance, correct reasoning from a set of beliefs seems to be quite constrained, despite the fact that most sets of beliefs are inconsistent. However, if explosion were valid, correct reasoning from an invalid set would presumably be completely unconstrained. Paraconsistent systems often have De Morgan negations. Philosophical debates about the acceptability of such logical systems often turn on the surprising cost of giving up explosion. Consider the following: If both P and $\neg P$ are true, we may choose an arbitrary Q and infer $(P \text{ or } Q)$ from P and then by so-called 'disjunctive syllogism' infer Q from $(P \text{ or } Q)$ and $\neg P$ – so we seem to have shown that explosion is a correct derived rule of inference. To block this argument, it seems we must: change our rule of 'or introduction'; hold that the result of stringing together deductively valid inferences is not always valid; or declare that disjunctive syllogism is not valid, despite appearances to the contrary. Many philosophers regard any of these as a higher price to pay than accepting the counterintuitive explosion rule.

Consider next the 'paradoxes of vagueness.' For concreteness, let $P(n)$ stand for 'there are more than n seconds in my childhood.' It seems we have good reason for accepting (i) $P(0)$ is true; (ii) for any n , if $P(n)$ is true then so too is $P(n + 1)$ (since one certainly does not change from a child to a nonchild from one second to the next); and (iii) for a sufficiently large m , $\neg P(m)$. However, $P(m)$ follows from (i) and (ii). In classical logic, rejecting (ii) implies that there is a number n such that one is a child for n seconds, but 1 second later one is not. However, in many logics this does not follow. In particular, many logics that reject the law of excluded middle avoid this conclusion. Indeed, the phenomenon of vagueness is thought by some to pose a more direct challenge to classical negation. The minimal logic condition applied to predicates instead of propositions means that $\neg P(x)$ picks out the property whose extension is as large as possible while remaining disjoint from the extension of $P(x)$. However, if $P(x)$ is vague, why should we suppose that all relevant x fall either in $P(x)$ or in $\neg P(x)$? To suppose there is always such a property seems to be assuming that every property has sharp

boundaries, which seems to amount to saying that there are no vague properties.

Finally, consider the ‘knowability paradox’: in classical logic, given some seemingly innocuous presuppositions about the nature of necessity and of knowledge, it follows from

(AR) all truths are *knowable*

that

(AK) *all truths are known.*

(Here is one set of assumptions suffices: knowledge distributes over conjunction; what is known is true; what is proved to be a theorem is thereby proved to be necessarily true; and if (necessarily $\neg P$) then \neg (possibly P).) Many philosophers are inclined to accept (AR), at least for certain domains of discourse, but almost none are tempted by (AK). However, the argument that (AR) implies (AK) depends essentially on double negation elimination. Hence, one can block the argument if negation is not De Morgan (e.g., if it is intuitionistic). Of course, to argue for this requires, if this solution is not to be merely *ad hoc*, giving independent reasons for rejecting classical negation – one could, after all, have rejected one of the plausible assumptions about necessity or knowledge instead, if blocking the paradox were alone sufficient to motivate such reformation of philosophical presuppositions.

More Than One Negation?

Some authors argue that it is a mistake to think that the same operation is involved in forming the negation of sentences and properties, something we have assumed without question. For instance, arguably the logical behavior of ‘ x is not- P ’ differs from that of ‘it is not the case that x is P ’; for example, x will be not-white only if x is definitely not white (i.e., is not a boundary case of whiteness), whereas if x is a boundary case, then it is not the case that x is white, but if the same operator is at work in both claims they should be equivalent.

Nonclassical languages sometimes suffer from a lack of expressive power if there is only one negation, in the sense that that negation leaves us unable to express what we seem able to express in natural language. For instance, $\neg P$, using intuitionistic negation, says that P is, in some suitable sense, impossible. (In both minimal and intuitionistic logic, $\neg P$ is often defined as ‘if P , then absurdity results.’ To see why this makes sense, look again at the characterization of minimal negation in the first section and note that adding explosion merely makes the consequences of absurdity particularly dire.) One loses the link to impossibility in De Morgan (and so classical) negations. For instance, with De Morgan negation

one can infer from $\neg(P \text{ and } Q)$ that $(\neg P \text{ or } \neg Q)$, but it certainly does not follow from the impossibility of P and Q being jointly satisfied that one of P and Q must be impossible. This matters because although one cannot intuitionistically derive (AK) from (AR), one can derive, for any P ,

(*) $\neg(P \text{ and } \neg(\text{It is known by someone at some time that } P))$

which is okay because this means only that it is impossible for P to be true and at the same time it be impossible for anyone to know it, which should follow from AR. However, if intuitionistic negation is the only one available it is difficult to see how one might express the claim that ‘some claims are true but never known’ since, as we have seen, (*) says something quite different. The obvious response is to argue that some other negation, compatible with but distinct from intuitionistic negation, is involved in expressing it. (Note that this other negation must be nonclassical, for if P has both a classical and intuitionistic negation, they are the same proposition, viz. the weakest proposition incompatible with P , so classical and intuitionistic negation cannot exist as distinct negation operators for the same class of sentences.) Similarly, if the only negation is the dialectic negation, how can we express things like ‘ P is true, and is not also false’?

Negation and Denial

Why not say, “The problem with (L) is that it *does not express a proposition*, and so *fails to have a truth value*”? Since it fails to have a truth value, so too does its negation – (L) and its negation fall into a *truth value gap*.” One reason is that this seems not to work for the so-called ‘strengthened liar’ (SL):

(SL) SL is not true.

For if (SL) is neither true nor false, then it is not true, and so should be true.

Notice that in our discussion of (L) we suggested that if P falls into a truth value gap, so too does $\neg P$, but in discussing (SL) we assumed that if P falls into a gap, not- P is true. If what we have said in both cases were correct, ‘not’ in the (SL) case is expressing some different notion of negation than is expressed by the negation sign in the (L) case, and we seem to be stuck with an ambiguity in the meaning of ‘not.’ It has been suggested, though, that ‘not’ is pragmatically, rather than semantically, ambiguous: It is sometimes used to form negations, and at other times it is used to deny a claim. When we say that (SL) is not true, we are denying (SL) without thereby asserting its negation.

However, we only fall back into paradox if we assert the negation of (SL). What makes this proposal philosophically contentious is that the prevailing view, going back at least to Frege, is that denial is not a separate speech act but is equivalent to the assertion of the negation of the claim in question.

See also: Logic and Language: Philosophical Aspects; Negation: Semantic Aspects; Paradoxes, Semantic; Semantics–Pragmatics Boundary; Vagueness: Philosophical Aspects.

Bibliography

- Dunn J M (1993). 'Perp and star: two treatments of negation.' In Tomberlin J (ed.) *Philosophy of language and logic: philosophical perspectives*, vol. 7. Atascadero, CA: Ridgeview. 331–357.
- Gabbay D & Wansing H (eds.) (1999). *What is negation?* Dordrecht, The Netherlands: Kluwer Academic.
- Horn L R (1989). *A natural history of negation*. Chicago: University of Chicago Press.
- Wansing H (ed.) (1996). *Negation: a notion in focus*. Berlin: de Gruyter.

Negation: Semantic Aspects

R van der Sandt, Radboud University, Nijmegen, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Classical and Nonclassical Negation

In logic, negation is an operation that applies to a sentence (or more generally a formula), thus yielding a new sentence. The sentence thus obtained is called the negation (or in some authors the denial) of the sentence negated. The negation sign is a logical operator, which is written in front of the relevant formula. The negation of a sentence φ is – depending on the notation convention adopted – written as $\neg\varphi$, $\sim\varphi$, $-\varphi$, $\bar{\varphi}$ or $N\varphi$. In Frege (1879), we find a small vertical stroke attached to the lower side of the content stroke. In his notation, the judgment that φ is not the case thus comes out as $\vdash\neg\varphi$.

In a classical two-valued logic (which allows exactly two truth values and moreover takes for granted that every sentence has a truth value), there is only one possible semantic definition of negation: the negation operator is a function that maps truth to falsity and falsity to truth. We may, however, deviate from classical semantics in various ways. One way is to question the assumption that every sentence has a truth value, another way is to allow for more than two truth values. This comes down to rejecting the principle of bivalence. The first strategy gives rise to so-called partial logics, the second to multivalent logics. A partial logic thus allows a sentence to have no value and it is common to refer to this third possibility as a truth value gap. In multivalent logics, we may define several negation operators; what their truth tables look like will depend on the intended interpretation of the nonclassical values. In a three-valued logic, a sentence that is not true may either be false or have the third truth value. In this view, falsity is a subcategory of what is not true. An alternative

account takes the third value as a special way of being not false. Finally, we may see the third value as independent of both truth and falsity. We thus may interpret the third value as a special way in which a sentence may go wrong and different interpretations will generally give rise to different truth tables.

Nonclassical accounts of negation go back to Aristotle. Aristotle argued (in his famous 'sea battle' argument) that if the sentence *There will be a sea battle tomorrow* were now either true or false, we would be committed to fatalism. Lukasiewicz concluded that such contingent future tense sentences are neither true nor false and consequently rejected bivalence. He introduced a third truth value to be interpreted as indeterminate or contingent, which naturally yields a truth table that maps truth to falsity, falsity to truth, and the third value to the third value. The Russian logician Bochvar gave another motivation for three-valued logics, arguing that it allows us to avoid logical paradoxes. Other writers have given various linguistic arguments for such deviations from classical logic. Three cases that received quite a bit of attention in the linguistic literature are category mistakes, vagueness, and presupposition failure.

Examples of category mistakes are Carnap's *Caesar is a prime number* or Chomsky's *Colorless green ideas sleep furiously*. Such sentences do not ascribe properties to things they do not have (which would result in falsity), but properties they cannot possibly have, thus yielding absurdity. And, since this holds for both the unnegated sentence and its negation, it gives a truth table that maps the third value (interpreted as nonsensical) on itself. The problem is closely related to the problem of presupposition failure (see below).

Vague predicates comprise predicates such as *small*, *beautiful*, *long*, etc., that is predicates that apply to a thing or set of things to a certain extent. In some cases, it is simply unclear whether a predicate

applies to an object. An example is Austin's *France is hexagonal*. Vague terms admit borderline cases, cases in which it cannot be uniquely determined whether a concept falls under the term or not. Three-valued logics are not of much help in solving the problem, since they replace the classic dichotomy between true and false with a strict trichotomy between true, false, and indeterminate, which yields results that are equally counterintuitive as those that result from a bivalent approach.

The most popular accounts are the so-called fuzzy logics introduced in the mid 1960s (Zadeh, 1965) and the method of supervaluations (Van Fraassen, 1966). Fuzzy logics accept an infinite number of truth values (a real number in the interval $[0, 1]$). The idea is that the closer the value is to 1 the 'truer' the sentence is. The value of a negated sentence is taken to be 1 minus the value of the unnegated sentence. Assuming that vague sentences get their value somewhere in the middle of the interval, it follows that if a sentence is maximally vague, so is its negation. The same holds for conjunctions and disjunctions of vague sentences. Fuzzy logics have been objected to on a number of counts. Most importantly, people have objected to the artificiality of assigning precise numerical values to the degree to which an object is small or beautiful. An alternative account is found in the method of supervaluations. Supervaluations – which were originally developed to account for presupposition failure – differ from multivalent logics in that they accept only two truth values but allow truth-value gaps. Just as in fuzzy logics, the negation of a vague sentence will come out as vague; but in contrast to fuzzy logics, formulas that are classically valid (or contradictory) will always come out as true or false no matter what the values are of their atomic components. A comparison between the approaches and a linguistic application is found in Kamp (1975).

Some of the best-known deviations from classical logic originate from presupposition theory. For Frege (1892) and Strawson (1952), presuppositions act as preconditions a sentence should satisfy in order to be true. Strawson and many other authors took preservation under negation as the defining characteristic. Thus (1b) seems to suggest as strongly as its unnegated counterpart (1a) that (1c) is true.

- (1a) The king of France is bald.
- (1b) The king of France is not bald.
- (1c) France has a king.

This phenomenon is general and seems to hold for all other lexical items that are said to induce presuppositions:

- (2a) John quit smoking.
- (2b) John didn't quit smoking.
- (2c) John used to smoke.

Inspired by Strawson the so-called neo-Strawsonians turned preservation under negation into a definition of semantic presupposition:

- (3) A sentence φ presupposes ψ just in case
 - (i) φ entails ψ ; and
 - (ii) $\neg \varphi$ entails ψ .

If we assume that negation maps truth to falsity and vice versa and, if we also adopt the standard definition of semantic entailment (φ entails ψ iff for any model M if $[[\varphi]]^M = 1$ then $[[\psi]]^M = 1$), (3) is equivalent to

- (4) φ presupposes ψ just in case
 - (i) in any model in which $[[\varphi]]^M = 1$ it holds that $[[\psi]]^M = 1$; and
 - (ii) in any model in which $[[\varphi]]^M = 0$ it holds that $[[\psi]]^M = 0$.

The latter simply comes down to the idea that if a sentence φ presupposes a sentence ψ , φ cannot have one of the classical values true or false, unless ψ is true. In a classical two-valued logic, this would predict that sentences can only presuppose tautologies. This forces the theorist to either allow for truth-value gaps or alternatively to introduce a third value to account for presupposition failure. If we allow truth-value gaps, compositionality predicts that if the unnegated sentence has no value, its negated counterpart will not have a value either. In a three-valued logic, this is captured by adopting the following truth table for negation (* indicates the third value):

(5)

φ	$\neg \varphi$
1	0
0	1
*	*

In conjunction with definition (3), this correctly predicts that presuppositions are preserved under negation.

Russell pointed out that (1b) has a second reading according to which this sentence is true simply because there is no king of France. In his analysis, the difference between the two readings is cashed out in terms of scope. In addition to the 'standard' interpretation, which is obtained by giving negation narrow scope and thus yields the formula (6a), we obtain the second reading by giving the negation scope over the full formula as in (6b):

- (6a) $\exists x \forall y [[KF(x) \leftrightarrow x = y] \wedge \neg \text{Bald}(y)]$
- (6b) $\neg \exists x \forall y [[KF(x) \leftrightarrow x = y] \wedge \text{Bald}(y)]$

This retains classical two-valued negation and eliminates presuppositions. It also illustrates the idea, found in Frege, that negation can always be represented as the standard operator suitably placed in the logical representation.

Russell's theory eliminates presuppositions. However, the difference between a presupposition preserving and presupposition canceling interpretation can also be obtained in a three-valued logic by defining a second negation operator $\neg \varphi$ that maps the third value onto truth:

(7)

φ	$\neg \varphi$
1	0
0	1
*	1

We perceive two kinds on nontruth. Nontruth as plain falsity and nontruth as a result of presupposition failure. The operator given in (5) is known as the weak Kleene-negation or the internal Bochvar connective; the operator defined in (7) is the same as the external Bochvar negation (also called the denial operator).

Negation and Polarity

Natural language contains a class of expressions whose distribution is limited to contexts that feel negative. Such items are called negative polarity items (NPIs). NPIs come in various types and syntactic categories: quantifiers and quantifier phrases (*any*, *any book*), verbs (*bother to*), factives (*mind that*, *matter that*), adverbs (*ever*, *yet*), prepositions (*until*), modals (*need*), idioms (*give a damn*, *lift a finger*), and many more. Thus the a sentences below, where we find the NPIs *any* and *ever* in the scope of a negative expression, are fine, but the b sentences are unacceptable.

- (8a) Mary didn't find any coin.
 (8b) *Mary found any coin.
 (9a) No one ever squared the circle.
 (9b) *Somebody ever squared the circle.

In addition to NPIs, we find a class of positive polarity items (PPIs). Positive polarity items typically occur in non-negative contexts. Just like NPIs, they belong to various syntactic categories. Examples of PPIs are quantifiers like *each* and *some*, factives like *be delighted that*, adverbs like *already*, and a rich variety of other expressions. The following examples show that PPIs shun negation:

- (10a) Harry is already serving breakfast.
 (10b) *Harry is not already serving breakfast
 (11a) Everybody inspected each package.
 (11b) *Nobody inspected each package.

Contexts that license NPIs however, need not contain overt negative elements. Thus negative polarity elements may be allowed in, for example, conditionals and *before*-clauses.

- (12a) If you find any clue, you should report it to the police.
 (12b) Mary left before John* already/even arrived.

In addition, some types of contexts only allow certain subclasses of polarity items. The central problem is thus to give a general characterization to the classes of contexts that license NPIs and PPIs and to relate different subclasses of polarity elements to the contexts that license them. On the question of why natural language contains such items and in view of what semantic properties makes them they behave as they do, there is no agreement whatsoever. Though most authors assume that there is a general mechanism that governs all such items, they also assume that positive or negative polarity is at least in part a conventional feature that is given in the lexicon.

Klima (1964) called contexts that license NPIs, affective contexts. Affective contexts in Klima's sense are contexts that are either explicitly negative or contain an underlying negative element NEG. Related views are put forward by Baker (1970) and Linebarger (1980), who claim that NPIs are licensed by negative contexts but the negative element need not surface. Baker requires that the licensing context be either negative or entail a negative sentence. Linebarger states that a NPI may be licensed by an overt negative element or a negative implicature. Ladusaw (1979) was the first to propose a semantic characterization of negative context as a NPI licenser. According to Ladusaw, an expression acts as a trigger for NPIs if it denotes a downward entailing function. The central notion is downward entailment or downward monotonicity. The definition runs as follows:

A function is downward entailing (monotone decreasing, downward monotonic) iff for all X and Y: if $X \subseteq Y$, then $f(Y) \subseteq f(X)$.

We can test for downward monotonicity by checking whether a context validates inferences from sets to subsets. Given that $\|chickpeas\| \subseteq \|peas\|$, the inference from (13a) to (13b) holds:

- (13a) Mary doesn't eat peas.
 (13b) Mary doesn't eat chickpeas.

And this predicts that (14a), where we find the NPI *any* in a downward entailing environment, is fine, though (14b) is unacceptable.

- (14a) Mary doesn't eat any peas.
 (14b) *Mary does eat any peas.

The converse property is upward entailment or upward monotonicity.

A function f is upward entailing (upward monotonic) iff for all X and Y : if $X \subseteq Y$, then $f(X) \subseteq f(Y)$.

The inference from *Some Texans eat chickpeas* to *Some Texans eat peas* is valid, which shows that *Some Texans* is upward entailing.

Most authors agree that Ladusaw's generalization is basically right when taken as a necessary condition (e.g., Van der Wouden). In order to license an NPI, the context should have the property of downward monotonicity. Zwarts (1986), Hoeksema (1983), and van der Wouden (1997) show that the issue is more complicated, however. Not all NPIs are licensed in all downward monotonic contexts. As a consequence, they distinguish between weak and strong NPIs (Zwarts) or between weak, medium, and strong ones (Van der Wouden). Weak NPIs conform to Ladusaw's generalization, NPIs of medium strength and strong ones require contexts that satisfy additional algebraic properties. The picture that arises from this work is that polarity licensing of both NPIs and PPIs is not a homogeneous phenomenon.

There is an additional complication. According to the characterization given above, putting a PPI in a negative context yields ungrammaticality. However, such constructions often allow for a special interpretation that suggests that the corresponding non-negative sentence has been uttered before. Seuren (1985) calls this the 'echo' effect and takes it as an argument for the existence of a special presupposition-canceling negation operator. For Horn (1985, 1989), it is a diagnostic for a special use of negation he calls metalinguistic. Thus (15) is acceptable as a reaction to the utterance of *Mary is still in Paris*.

(15) Mary is NOT still in Paris. She never intended to be there and went to London instead.

It is implausible, though, that this phenomenon has anything to do with the formal properties of negation or negative context. We observe the same effect when NPIs are put in uncontroversially positive contexts.

(16) It DOES matter that my bunny is ill.

We return to the phenomenon of polarity reversal at the end of the last section.

Negation Versus Denial

The negation of a sentence is sometimes called its denial. It is, however, important to strictly separate the concepts of denial and negation. The concept of denial belongs to speech act theory. In this respect,

denials are on a par with assertions. And, in contrast to negation, denials – just like assertions – should not be characterized in terms of truth and falsity but in functional terms. The primary function of assertions is to introduce new information, that is information that is not already taken for granted by the participants in a discourse. Denials differ in this respect. Their essential function is to object to information that has already been introduced before or is in some sense taken for granted. In doing so, a denial will have the effect of removing information from the discourse record. Despite being overtly negative, (17a) and (17b) will, if processed as the first sentence of a discourse, not be interpreted as denials but as assertions of negative sentences.

(17a) Mary is unhappy/not happy.

(17b) It does not matter that Mary read your letters.

Note, moreover, that a denial need not contain a negative morpheme. If uttered as a reaction to (17a) or (17b), B-b and B-c below will not be interpreted as assertions but as denials of the utterances they object to. The utterance by means of which the denial is performed may thus be of a negative or positive form. This shows that the concept of denial is logically independent of the concept of negation.

The difference in status of the concepts of denial and negation gave rise to a view according to which denials can be semantically characterized in terms of assertion and negation. Following Frege–Austin, Searle, and Dummett claimed that denials can be semantically analyzed as assertions of negative sentences. Further pragmatic effects that distinguish denials from assertoric utterances must be accounted for independently. Frege (1918: 149; Geach 1997: 40) – states: "People speak of affirmative and negative judgments; even Kant does so [...] For logic at any rate such a distinction is wholly unnecessary; its grounds must be sought outside logic." A linguist like Givón (1978) implicitly adopts the view that negation in language should be analyzed as a logical sign, which operates on sentences, but points out that denials have, in addition to their logical aspect, a pragmatic function that should be treated independently. Thus, though denials should logically be treated as negations of positive sentences, they constitute a different speech act from assertions. "While the latter are used to convey new information on the presumption of ignorance of the hearer, negatives are used to correct misguided belief on the assumption of the hearer's error." (Givón, 1978).

However, Frege also points out that it is not easy to make sense of the interpretation of denials as negative statements. "It is by no means easy to state what is a

negative judgment (thought). Consider the sentences ‘Christ is immortal’, ‘Christ lives for ever’, ‘Christ is not immortal’, ‘Christ does not live for ever’. Now which of the thoughts here is affirmative, which negative?ⁿ (Frege, 1918: 150; Geach 1977: 41). The central problem is that a treatment of the negation sign as a sign of denial forces us to introduce a further operator to account for negatory or denial force, in addition to the negation operator. Clearly, no operator that would go or fuse with Frege’s assertion sign can be given a sensible interpretation in an embedded environment. Though an interpretation could be given to such an operator for simple sentences in an unembedded environment, it precludes the possibility of using this very same sentence as the antecedent of a conditional sentence. In this context, the sentence cannot have denial of negatory force. Fregean force always goes with full sentences and does not contribute to the proposition expressed. Thus in embedded environments, negation can only be interpreted as an ordinary functional expression, reversing the truth value of the proposition expressed. However, the introduction of some kind of denial operator having the same role as assertion would force us to recognize two ways of judging and accordingly two negative operators of a different status, one as some kind of speech act device, the other as a semantic operator reversing the truth value of the proposition expressed. Consequently, we should separate force from content, thus ending up with a theory that is preferable for being both conceptually and formally simpler.

Metalinguistic Negation

Horn (1985, 1989) drew attention to the fact that an utterance may not just be rejected in view of falsity of the propositional content, but that denials may instead apply to a variety of information, comprising presuppositions, various types of implicatures and connotations related to style and register. He points out that denials can be used to reject an utterance of a previous speaker for whatever reason. And equating the notions of denial and negation, he made a distinction between the well-known truth-functional operator and a non-truth-functional metalinguistic device. On this view, standard truth-functional negation is found in negative assertions and propositional denials. His metalinguistic device applies to presuppositional inferences, implicatures, and connotations of style and register. He labeled this device ‘metalinguistic negation’ and characterized it as follows: “I am claiming [...] for negation [...] a use distinction: it can be a descriptive truth-functional operator, taking

a proposition p into a proposition $\neg p$, or a metalinguistic operator which can be glossed ‘I object to u ’, where u is crucially a linguistic utterance rather than an abstract proposition” (Horn, 1985: 136). The relevant examples can be classified in the following categories:

- A Assertions of negative sentences:
 - (A-a) Mary is unhappy.
 - (A-b) Mary is not happy.
 - (A-c) It does not matter that Mary read your letters.
- B Propositional denials:
 - (B-a) Mary is NOT happy (as a reaction to the utterance of ‘Mary is happy’).
 - (B-b) Mary is happy (as a reaction to the utterance of A-a or A-b).
 - (B-c) It DOES matter that Mary read my letters.
- C Presuppositional denials:
 - (C-a) The king of France is NOT bald – France does not have a king.
 - (C-b) Virginia CANNOT know that the earth is flat.
 - (C-c) John DIDN’T stop smoking – he never smoked.
- D Implicature denials:
 - (D-a) It is not POSSIBLE, it is NECESSARY that the church is right.
 - (D-b) That haggis is not GOOD, it is EXCELLENT.
 - (D-c) That was not a LADY I kissed last night – it was my WIFE.
- E A variety of connotations (conventional implicature, style, register)
 - (E-a) That is not a STEED – it’s a HORSE.
 - (E-b) Grandma didn’t KICK THE BUCKET – she passed away.
 - (E-c) He didn’t call the police, he called the POLICE.

Abstracting from their discourse function and focusing on truth conditional content, we may first note that Frege’s strategy of analyzing the denial B-a as an assertion of a positive sentence works as far as truth conditional content goes. Assuming double negation in their underlying form, B-b and B-c can also be analyzed in the Fregean way. The presuppositional denials under C have historically either been handled by means of some nonstandard logic (thus postulating an ambiguity in the negation operator), or by an appeal to the Russellian notion of scope (see the first section above). More recent accounts that are formulated in a dynamic framework achieve a similar effect by invoking the notion of local accommodation.

Though such theories take it that presuppositions normally escape from the scope of negation and other embedding operators, they introduce a special mechanism that forces presuppositional information – under specified conditions (e.g., threatening inconsistency) – to remain *in situ*, i.e., the place where they were triggered, and thus keep them within the scope of the embedding operator (Heim, 1983; Van der Sandt, 1992). The examples under D and E are problematic for any purely logical account. Their standard translations amount to plain contradiction, which led authors to call for some non-truth-functional analysis or else a nonliteral reinterpretation in view of a violation of Gricean maxims.

Horn treats categories C–E as metalinguistic uses of negation. The following diagnostics act as a kind of litmus test:

- Metalinguistic negation requires wide scope, thereby blocking lexical incorporation.
- Such sentences exhibit a rising intonation contour (fall-rise) pending further clarification, thereby stressing the offensive item or else triggering contrastive stress on the negative morpheme.
- Metalinguistic negation reverses the polarity of the sentence.

Example (18) illustrates the first diagnostic. In contrast to example D-a, it amounts to a straightforward contradiction:

- (18) It is impossible, it is necessary that the church is right.

Note, moreover, that the same holds for example D-a if we destress the offensive item *possible*. Sentences (15) and (16) show that – contrary to received wisdom – negative denials accept PPIs, while positive denials allow NPIs.

Van der Sandt (1991, 2003) stressed that in view of their function, denials should be interpreted in context. They are not incremental in the sense assertions are, nor can they naturally occur in isolation. They take up or ‘echo’ a previous utterance; and since utterances convey information of various kinds (comprising presuppositions, implicatures, and further connotations), the interpretation of a denial may be sensitive to all the information that is invoked by the utterance they object to. Just as a speaker, when uttering any of the B sentences, objects to the truth-conditional content of a previously uttered sentence, he objects in the C cases to the presuppositions carried by a previous utterance, in the D cases to implicatures invoked, and in the E cases to offensive connotations of style and register a previous speaker

may have conveyed with a corresponding non-negative utterance. Consider (19b), which is typically a reaction to (19a)

- (19a) It is possible that the church is right.
 (19b) It is not POSSIBLE, it is necessary that the church is right.

The utterance of (19a) invokes the implicature that it is not necessary that the church is right. By denying this utterance by means of (19b), the second speaker does not just react to the propositional content but to all the information conveyed by (19a). This information comprises the (scalar) implicature that it is not necessary that the church is right. Thus, if we take the negation in (19b) to apply to the full informative content of the first utterance, this yields an interpretation that can be paraphrased as (20):

- (20) $\neg(\Diamond \text{the_church_is_right} \wedge$
 $\neg \Box \text{the_church_is_right}) \wedge$
 $\Box \text{the_church_is_right}$

This simply conveys that it is necessary that the church is right and rejects all information conveyed by the previous utterance. The formal account is implemented in a nonincremental theory of discourse processing.

While Horn subsumes only presuppositional denials, implicature denials and the style and register cases under his notion of metalinguistic negation, Van der Sandt takes the phenomenon to be fully independent of the notion of negation and proposes an explanation of both the Horn cases and propositional denials based on the discourse properties of denial. Geurts (1998) objects to both views. According to him, denial or metalinguistic negation is not a homogeneous phenomenon. Instead he claims that a different semantic mechanism is operative in each of the cases B–E. His account divorces metalinguistic negations from the discourse characteristics of denial. Propositional denials are simply assertions of negative sentences. Presuppositional denials are handled by the mechanism of local accommodation. For the last two categories, Geurts assumes that the negative morpheme operates on a different level. For implicature denials, he invokes the notion of semantic polysemy. Crucially, he maintains that a scalar like *possible* may either have an *at least* or an exhaustive sense (i.e., *possible and not necessary*). Example (19a) selects the *at least* interpretation and here the exhaustive interpretation comes about by way of conversational implicature. In the denial (19b), the exhaustive sense is selected immediately. Only the style and register and pronunciation cases are metalinguistic in that they make reference to linguistic objects. Example

E-b typically corrects the pronunciation of *police*. In this context, the phrase POLice corrects pronunciation, and here this phrase may be constructed as ‘the body whose name is pronounced “police”.’

Returning to the second section of this article, I conclude with a remark on polarity reversal. On the account that strictly distinguishes between the use of a (negative) sentence to convey new information and its use as a denial, polarity reversal falls out as a consequence of the discourse function of denial.

Consider first the unmarked (21a) and (21b):

(21a) It does not matter that Mary has read my letters.

(21b) Harry did not pick any of the flowers.

A denial interpretation of either of these sentences would require that the corresponding non-negative sentence has been uttered just before (or is at least ‘in the air’). Note that their non-negated counterparts would contain a negative polarity element in a non-negative context and thus be ungrammatical. This forces the regular assertoric interpretation. But (16), which does contain an NPI in a non-negative context, can very well be used to echo a negative and thus grammatical utterance. An analogous explanation can be given for negative sentences containing PPIs. Since affirmative sentences containing PPIs require a non-negative context, their negated counterparts cannot be interpreted as isolated assertions. But they can very well be interpreted as the denial of a previous utterance, which being positive, is fully grammatical. Sentence (15) illustrates this.

See also: Negation: Philosophical Aspects; Presupposition.

Bibliography

- Baker C L (1970). ‘Double negatives.’ *Linguistic Inquiry* 1, 169–186.
- Dummett M (1973). *Frege. Philosophy of language*. London: Duckworth.
- Frege G (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Louis Nebert. English translation in Geach P & Black M (1960). *Gottlob Frege. Logical investigations*. Oxford: Blackwell. 1–20.
- Frege G (1892). ‘Über Sinn und Bedeutung.’ *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50. English translation in Geach & Black, 56–78.
- Frege G (1918). ‘Die Verneinung.’ *Beiträge zur Philosophie des deutschen Idealismus* 1, 143–157. English translation in Geach P (ed.) (1977) *Gottlob Frege. Logical investigations* Oxford: Blackwell. 30–53.
- Geach P & Black M (1960). *Translations from the philosophical writings of Gottlob Frege* (2nd edn.). Oxford: Blackwell.
- Geurts B (1998). ‘The mechanisms of denial.’ *Language* 74, 274–307.
- Givón T (1978). ‘Negation in language: pragmatics, function, ontology.’ In Cole P (ed.) *Syntax and semantics 9: pragmatics*. New York: Academic Press. 69–112.
- Heim I (1983). ‘On the projection problem for presuppositions.’ In *Proceedings of the West Coast Conference in Formal Linguistics (WCCFL)* 2. 114–126.
- Hoeksema J (1983). ‘Negative polarity and the comparative.’ *Natural Language and Linguistic Theory* 1, 403–434.
- Horn L R (1985). ‘Metalinguistic negation and pragmatic ambiguity.’ *Language* 61, 121–174.
- Horn L R (1989). *A natural history of negation*. Chicago: University of Chicago Press.
- Kamp H (1975). ‘Two theories about adjectives.’ In Keenan E (ed.) *Formal semantics for natural language*. London: Cambridge University Press. 123–155.
- Klima E S (1964). ‘Negation in English.’ In Fodor J & Katz J J (eds.) *The structure of language*. Englewood Cliffs: Prentice Hall. 246–323.
- Ladusaw W A (1979). Polarity sensitivity as inherent scope relations. Ph.D. diss. University of Texas at Austin. Distributed by IULC, Bloomington, Indiana (1980).
- Linebarger M (1980). ‘The grammar of negative polarity.’ Ph.D. diss. MIT. Distributed by IULC, Bloomington, Indiana.
- Seuren P A M (1985). *Discourse semantics*. Oxford: Blackwell.
- Strawson P (1952). *Introduction to logical theory*. London: Methuen.
- Van der Sandt R A (1991). ‘Denial.’ In *Papers from Chicago Linguistic Society* 27(2): the parasession on negation. 331–344.
- Van der Sandt R A (1992). ‘Presupposition projection as anaphora resolution.’ *Journal of Semantics* 9, 333–377.
- Van der Sandt R A (2003). ‘Denial and presupposition.’ In Kühnlein P & Zeevat H (eds.) *Perspectives on dialogue in the new millennium*. Amsterdam: John Benjamins.
- Van der Wouden T (1997). *Negative contexts. Collocation, polarity and multiple negation*. London: Routledge.
- Van Fraassen B (1966). Singular terms, truth value gaps and free logic. *Journal of Philosophy* 63, 481–495.
- Zadeh L (1965). ‘Fuzzy sets.’ *Information and Control* 8, 338–353.
- Zwarts F (1986). ‘Categoriale grammatika en algebraïsche semantiek. Een Studie naar Negatie en Polariteit in het Nederlands.’ Ph.D. diss. Groningen University.

Nominalism

G Klima, Fordham University, Bronx, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

Nominalism is usually characterized in contrast to other two major theoretical alternatives concerning the ontological status of universals. According to this characterization, the nominalist position holds that universals are mere words (e.g., the word ‘round’); in contrast, the conceptualist position would identify universals with concepts (e.g., the concept expressed both by the English word ‘round’ and by the Latin word ‘rotundus’), and the realist position would claim universals to be universal things (e.g., the nonmental, nonphysical, abstract Form of Roundness itself).

Both historically and theoretically, there are a number of problems with this simple, indeed, simplistic conception. First, by these standards, no premodern author would qualify as a nominalist since all these authors held that our universal terms owe their universality to the universal concepts they express (thus, on this account all these authors would have to be classified as conceptualists), and no medieval author would count as a realist because they all denied the real existence of Platonic ‘abstract entities’ (although most medieval authors did posit universals both *in re*, as trope-like individualized forms such as the individual roundness of *this* billiard ball as opposed to the numerically distinct roundness of another one, and *ante rem*, as Divine Ideas, identical with God, who is not an ‘abstract entity’) (Klima, 2001). Accordingly, this conception leaves in obscurity the genuine theoretical differences of medieval nominalists and realists, who first distinguished themselves under these designations. However, as we shall see, clarifying these genuine theoretical differences, we can have a deeper understanding not only of what a genuinely nominalist position was in the late Middle Ages but also what in general a genuinely nominalist position has to consist in, as well as the important conceptual connections such a position has to such broader issues as ontological commitment, abstraction, induction, essentialism, and the possibility of valid scientific generalizations.

Extreme Realism: Plato’s Ideal Exemplars

The problem of universals originated with Plato’s answer to the question of how universal knowledge (e.g., our knowledge of geometrical theorems) of a potential infinity of individuals is possible. Plato’s answer in terms of his theory of Forms involves the idea of regarding individuals of the same kind as

copies of an original exemplar or archetype and assuming that our understanding has some direct access to this exemplar. Clearly, if I can read today’s news on the printer’s printing plates, then I do not have to buy the copies at the newsstands to get my news. Likewise, knowing that the form of all triangles, after which all triangles are modeled, has three angles equal to two right angles, I know that all triangles have the same property. Thus, Plato’s answer is obviously a good answer to the question of the possibility of universal knowledge. However, this answer raises more problems than it solves. If the exemplar is not one of the copies, yet it has to be like the copies (given that the copies all have to be like it), then what sort of entity is it? In fact, this is precisely the basis of the most famous argument against Plato’s theory, the Third Man (Wedberg, 1978). For if the exemplar has to be of the same sort as its copies, it can be grouped together with its copies into the same kind. However, if for a group of individuals of the same kind there always has to be a common exemplar, and nothing can be its own exemplar, as Plato’s theory claims, then there has to be another exemplar for the exemplar and its copies taken together. Since this reasoning can be repeated indefinitely, we have to conclude that for a given set of copies there would have to be an infinity of exemplars, contrary to the theory’s explicit claim that there can only be one such exemplar. Therefore, since the theory as stated entails inconsistent claims, it cannot be true in this form.

Moderate Realism or Conceptualism: Aristotle’s Universals

Prompted by such and similar inconsistencies, Aristotle famously rejected Plato’s Forms and provided a radically different sort of answer to Plato’s original question concerning the possibility of universal knowledge. Plato’s idea was that this sort of knowledge is possible because of the human soul’s prenatal access to the Forms serving as the common exemplars of all sorts of particulars the soul comes to experience in this life and the soul’s ability to recollect these Forms and their properties prompted by these experiences. In contrast, Aristotle’s idea was that we can have this sort of knowledge from experience in this life. To be sure, the finite number of experiences we can have in this life with individuals of the same sort can never justify universal knowledge claims that cover a potential infinity of such individuals. However, if our cognitive faculties enable us to recognize a common pattern that equally

characterizes any possible individual of the same sort, then even this finite number of experiences may enable us to make valid generalizations covering all possible individuals of the same sort. Using the previous analogy, in this case I do not have access to the printer's plates (if there is any such thing at all, which need not be the case if copies can 'self-reproduce,' as in photocopying); I can only collate a number of variously smudged, incomplete newspapers, from which, nevertheless, I am able to extract their common content so that I will know the news any other copy of the same issue may carry even before actually checking it. This is the idea Aristotle worked out in his theories of abstraction and induction.

The Moderate Realism/Conceptualism of Medieval Aristotelians

According to Aristotle and his medieval interpreters, human understanding is characterized by two intellectual powers or faculties, namely the active intellect (*nous poietikos*, *intellectus agens*) and the receptive intellect (*nous pathetikos*, *intellectus possibilis*). Due to the obscurity of Aristotle's formulations, in the Middle Ages there was much debate about the makeup of these faculties, namely whether they are material or immaterial, and whether they existed as individualized powers of individual human souls or, rather, as separate substances connected to individual human souls to carry out their intellectual operations (pretty much as mainframe computers are connected to their terminals). However, regardless of these differences, there was quite general agreement concerning the function of these intellectual powers (whatever they are in themselves) and how they can produce universal knowledge without having to 'look up to' or 'recollect' Platonic Forms. The active intellect has the function of producing simple, universal concepts (i.e., universal representations of individuals) from their singular representations provided by the senses. The receptive intellect, on the other hand, receives these universal representations for further processing, combining them into more complex concepts and judgments by means of complexive concepts, the so-called syncategorematic concepts ('syncategoremata'), and using these judgments in reasoning.

Abstraction, Induction, and Essentialism

In the process of abstraction, the active intellect forms a common concept by considering the multitude of experiences of particulars of the same kind, disregarding what is peculiar to each while focusing on what is common to all. The common concept retained

in the receptive intellect is a mental representation that represents the individuals it is abstracted from only with respect to what commonly characterizes them all. Accordingly, the same mental representation will naturally represent in the same respect not only those individuals that it is abstracted from but also other individuals of the same kind – that is, individuals that resemble the observed ones precisely in the same respect in which the abstracted concept represents the observed individuals. Therefore, if the receptive intellect forms a judgment with this concept in which the predicate belongs to the particulars that fall under this concept on account of falling under this concept, then it is able to produce a valid scientific generalization by induction – that is, by reasoning in the following way: This observed S was P, and that observed S was P, and so on for all observed S's; therefore, in general, every S is P. Obviously, this inference is valid (not formally, but materially) only if an S is essentially and not merely coincidentally P – that is, if an S is P on account of being an S so that any S is necessarily P, as long as it is an S (and if it is also essentially an S, then it is necessarily an S, and hence also a P, as long as it exists). This is precisely why this inference requires the multitude of experiences, namely to ensure that the observed S's have been P not by mere coincidence but precisely on account of being S's. However, if this inference is valid, then it obviously provides a good answer to Plato's original question (concerning the possibility of universal knowledge) without resorting to Plato's universal Forms.

The Ontological Commitments of Moderate Realism

However, the previous conception may still involve an enormous amount of ontological commitment; admittedly, not to Platonic universal Forms but to individuals of a rather strange sort, the individualized forms of individual substances. If all our common terms are applicable to individual things on account of our common concepts, and we gain these concepts by abstraction from the individualized forms or properties that sort individuals into various kinds, then apparently there are as many individualized forms in an individual substance as there are true common predicates of it. (Yet, this conception need not make individual substances into mere congeries of individualized forms, in the vein of modern trope theories. On this conception, individual substances are the primary units of reality that account for the individuation of individualized forms, and not the other way around (cf. Campbell, 1997; Daily, 1997).) In this moderate realist view, commonly endorsed

by medieval authors of the 13th century (e.g., Thomas Aquinas), the world is a world of individuals without ‘universal entities,’ but it is populated by all sorts of nonsubstantial particulars, serving as the individualized *significata* of our common terms. The universals of this world, then, are the common terms of our language as well as the concepts they express, and the individualized forms as-conceived-in-a-universal-manner by means of these concepts (i.e., as-existing-in-the-mind as the direct objects of these concepts) (Klima, 1999). To be sure, these authors were able to reduce the ontological commitment of their theories by two fundamental semantic strategies: the distinction of several senses of ‘being’ and the identification of the semantic values of various terms.

On the first strategy, ontological commitment simply becomes ambiguous among different kinds of entities credited with different degrees of reality corresponding to different senses of ‘existence’ or ‘being’ (such as being in reality, *esse reale*, and being in the mind, *esse rationis*; cf. Klima, 1993) – a move generally rejected by nominalists.

On the second strategy, however, ontological commitment is reduced even within the same domain of entities. For instance, since two billiard balls A and B are similar in shape, the moderate realist view is apparently committed not only to the individualized roundness of A and the individualized roundness of B but also to the individualized similarity of A to B and to the individualized similarity of B to A. However, several moderate realist authors would argue that nothing prevents us from saying that the roundness of A is the same thing as the similarity of A to B, merely conceived differently, by means of a different concept (Henninger, 1989; Brower, 2001).

Late Medieval and Modern Nominalism

This strategy actually anticipates the more radical reductionist program of William Ockham and his 14th-century followers, such as John Buridan, Albert of Saxony, Marsilius of Inghen, and in general the late medieval nominalist tradition of the *via moderna*. These nominalist authors agreed with their moderate realist counterparts in positing a world consisting only of individuals and identifying universals with universal terms of written and spoken languages, which owe their universal mode of representation to universal concepts of the human mind, interpreted as singular mental acts representing their singular objects in a universal manner. However, they denied even the ‘diminished’ existence of universals as universal objects of these mental concepts admitted by moderate realists, and they reduced the number of

really distinct categories of singulars to two or three, admitting distinct entities only in the categories of substance, quality, and possibly quantity. (Ockham, for example, identified substance with quantity, whereas Buridan argued against their identification.)

To be sure, using their own eliminative strategy of identifying the semantic values of linguistic items in different linguistic categories, the moderate realists could in principle achieve the same degree of ontological parsimony. However, for them this was to be achieved separately, based on metaphysical considerations. Their semantics demanded positing as many different types of semantic values for their terms as there are different linguistic categories, although it was possible for them to decide on the basis of further metaphysical considerations about the identity or nonidentity of these semantic values. In contrast, the nominalists had the principles of ontological reduction (i.e., eliminating unwanted ontological commitment) ‘built into’ their semantic principles (Klima, 1999). The reductionist program of the nominalists was devised to show that it is possible to have a sufficiently ‘fine-grained’ semantics without a complex ontology by ‘moving’ the requisite distinctions from (a full-fledged or ‘diminished’) reality to the conceptual structures posited in the mind.

For example, in contrast to the moderate realist interpretation of similarity alluded to previously, in the nominalist approach there is no metaphysical question of whether the relation signified by the term ‘similar’ is identical with its foundation, for example, the roundness of ball A. The signification of ‘similar’ is not construed in the same way in the first place. Instead of assuming that in ‘A is similar to B’ this term signifies A’s similarity to B, an extramental entity that may or may not be distinct from the roundness of A, the nominalists would say that this term simply signifies A in relation to B, on account of the relative (connotative) concept whereby we conceive of A in relation to B. Therefore, even if we do have the obvious semantic difference between the absolute term ‘round’ and the relative term ‘similar,’ this semantic difference need not be accounted for in terms of a bloated ontology of distinct extralinguistic correlates of these terms because the relevant semantic distinctions can be made with reference to conceptual distinctions in the mind.

Following this nominalist strategy, once we have identified a basic ‘vocabulary’ of simple concepts that only commit us to entities in the ‘permitted’ ontological categories, any further apparent ontological commitments of our language can be eliminated by means of nominal definitions, which show how the semantic features of linguistic items carrying such apparent ontological commitment can be

accounted for in terms of an implicit conceptual structure that can be explicated by means of the basic vocabulary carrying commitment only to permitted entities. Indeed, this strategy of eliminating ontological commitment by means of nominal definitions explicating this implicit conceptual structure foreshadows the similar strategy of elimination by paraphrase characteristic of all modern nominalistic programs.

Thus, in this approach, and this is the gist of medieval nominalism in general, a sufficiently fine-grained semantics for natural language is achieved by mapping linguistic constructions onto sufficiently rich conceptual structures in a mental language (comparable in its role to the contemporary 'language of thought hypothesis'), which in turn can be mapped onto a parsimoniously construed reality without any loss of semantic distinctiveness.

Modern nominalistic programs, such as the program promulgated by Goodman and Quine (1947), apply the same basic strategy, but without the medievals' appeal to a mental language, by introducing the explicit ontological commitment of a primitive vocabulary and by eliminating any further apparent commitment (e.g., to such 'abstract entities' as numbers) by providing suitable paraphrases of relevant linguistic items (e.g., numerals) in terms of this primitive vocabulary. Indeed, in this program, Platonistic descriptions of language in terms of linguistic types are replaced by a nominalistic syntax treating of tokens, such as singular inscriptions. (In fact, medieval nominalists also formulated their logical theories in terms of token-phrases (Klima, 2004a).) The successful paraphrase then shows that there is no need to assume the existence of the putative entities apparently required by the semantic features of the phrases that appeared to carry commitment to them, so nominalists are entitled to get rid of them by one swoosh of Ockham's razor.

Nominalism, Antirealism, and Skepticism

Yet, despite customary charges and modern tendencies to the contrary, the 'reductionist' program and the corresponding strategy of medieval nominalism did not necessarily result in metaphysical anti-realism, conventionalism, or skepticism. Medieval nominalists typically regarded concepts as naturally representative of a world of individuals presorted into natural kinds and not sorted into these kinds by our concepts and/or linguistic conventions. Thus, they maintained an essentialist metaphysics, and so the scientific knowability of a mind-independent reality, yet without any ontological commitment to (whether

subsistent or inherent) universal essences distinct from their individuals. Whether they could consistently do so is a further issue, which bears direct relevance to contemporary considerations concerning ontological commitment and metaphysical essentialism. Plato's problem of the possibility of universal knowledge still remains a problem for nominalists: Once they have rejected the moderate realists' distinct individualized forms, which in the moderate realist conception are precisely the items in reality that serve as the foundation for the abstraction of our universal concepts, the nominalists still have to provide a plausible story about how universal concepts can be abstracted from observed singulars (and hence how they can apply even to previously unobserved singulars), unless they want to give up on the possibility of universal knowledge in the classical sense altogether (Klima, 2004b, 2005).

See also: Concepts; Empiricism; Mentalese; Natural Kind Terms; Objects, Properties, and Functions; Semantic Value; Syncategoremata.

Bibliography

- Aydede M (2004, Fall). 'The language of thought hypothesis.' In Zalta E N (ed.).
- Bacon J (2002, Fall). 'Tropes.' In Zalta E N (ed.).
- Balaguer M (2004, Summer). 'Platonism in metaphysics.' In Zalta E N (ed.).
- Brower J (2001, Summer). 'Medieval theories of relations.' In Zalta E N (ed.).
- Campbell K (1997). 'The metaphysic of abstract particulars.' In Mellor & Oliver (eds.). 125–139.
- Daily C (1997). 'Tropes.' In Mellor & Oliver (eds.). 130–159.
- Goodman N & Quine W V O (1947). 'Steps toward a constructive nominalism.' *Journal of Symbolic Logic* 12, 105–122.
- Henninger M (1989). *Relations: medieval theories 1250–1325*. Oxford: Clarendon.
- Klima G (1993). 'The changing role of *Entia Rationis* in medieval philosophy: a comparative study with a reconstruction.' *Synthese* 96, 25–59.
- Klima G (1999). 'Ockham's semantics and metaphysics of the categories.' In Spade P V (ed.) *The Cambridge companion to Ockham*. Cambridge, UK: Cambridge University Press. 118–142.
- Klima G (2001, Winter). 'The medieval problem of universals.' In Zalta E N (ed.).
- Klima G (2004a). 'Consequences of a closed, token-based semantics: the case of John Buridan.' *History and Philosophy of Logic* 25, 95–110.
- Klima G (2004b). 'John Buridan on the acquisition of simple substantial concepts.' In Freidmann R L & Ebbesen S (eds.) *John Buridan and beyond: topics in the language*

- sciences 1300–1700*. Copenhagen: The Royal Danish Academy of Sciences and Letters. 17–32.
- Klima G (2005). ‘The essentialist nominalism of John Buridan.’ *Review of Metaphysics* 58, 301–315.
- Lagerlund H (2004, Summer). ‘Mental representation in medieval philosophy.’ In Zalta E N (ed.).
- McInerney R (2002, Spring). ‘Saint Thomas Aquinas.’ In Zalta E N (ed.).

- Mellor D H & Oliver A (eds.) (1997). *Properties*. Oxford: Oxford University Press.
- Rosen G (2001, Fall). ‘Abstract objects.’ In Zalta E N (ed.).
- Wedberg A (1978). ‘The theory of ideas.’ In Vlastos G (ed.) *Plato: a collection of critical essays*. Notre Dame, IN: University of Notre Dame Press. 28–52.
- Zalta E N (ed.) (2001–2004). *Stanford encyclopedia of philosophy*. <http://plato.stanford.edu>.

Nonmonotonic Inference

K Frankish, The Open University, Milton Keynes, UK

© 2006 Elsevier Ltd. All rights reserved.

In most logical systems, inferences cannot be invalidated simply by the addition of new premises. If an inference can be drawn from a set of premises *S*, then it can also be drawn from any larger set incorporating *S*. The truth of the original premises guarantees the truth of the inferred conclusion, and the addition of extra premises cannot undermine it. This property is known as *monotonicity*. (The term is a mathematical one; a monotonic sequence is one whose terms increase but never decrease, or vice versa.) *Nonmonotonic* inference lacks this property. The conclusions drawn are provisional, and new information may lead to the withdrawal of a previous conclusion, even though none of the original premises is retracted.

Much of our everyday reasoning is nonmonotonic. We frequently jump to conclusions on the basis of partial information, relying on rough generalizations – that people usually mean what they say, that machines usually work as they are designed to, that objects usually stay where they are put, and so on. We treat these conclusions as provisional, however, and are prepared to retract them if we learn that the cases we are dealing with are atypical. To take an example that is ubiquitous in the literature, if we know that Tweety is a bird, then we may infer that Tweety flies, since we know that birds typically fly, but we shall withdraw this conclusion if we learn that Tweety is an atypical bird – a penguin, say.

An important feature of inferences like this is that they are sensitive to the absence of information as well as to its presence. Because we lack information that Tweety is atypical, we assume that he is not and proceed on that basis. (That, of course, is why the acquisition of new information can undermine the inference.) In standard logics, by contrast, inference is sensitive only to information that is explicitly represented, and we would need to add the premise that Tweety is not atypical in order to reach the conclusion

that he flies. This feature of nonmonotonic inference makes it highly useful. We do not have the time or mental capacity to collect, evaluate, and process all the potentially relevant information before deciding what to do or think. (Think of everything we would need to know in order to be sure that Tweety is not atypical – that he is not a penguin, not an ostrich, not a hatchling, not injured, not tethered to the ground, and so on.)

Because of its central role in commonsense reasoning, nonmonotonic inference has attracted much interest from researchers in artificial intelligence – in particular from those seeking to model human intelligence in computational terms. The challenge has been to *formalize* nonmonotonic inference – to describe it in terms of a precisely defined logical system that could then be used to develop computer programs that replicate everyday reasoning.

Nonmonotonic logic also has applications to more specific problems in artificial intelligence, among them the so-called *frame problem* (McCarthy and Hayes, 1969; see also **Frame Problem**). In order to plan how to reach its goals, an artificial agent will need to know what will and what will not change as a result of each action it might perform. But the things that will *not* change as a result of an action will be very numerous, and it would be impracticable to list them all in the system’s database. A more efficient solution would be for the system to reason nonmonotonically, using the rule of thumb that actions leave the world unchanged except in those respects in which they are known to alter it. Another application is in the area of database theory. Here it is often convenient to operate with the tacit assumption that a database contains all the relevant information and to treat as false any proposition that cannot be proved from it. This is known as the *closed world assumption* (Reiter, 1978). Again, this involves a nonmonotonic inference relation, since the addition of new data may permit the derivation of a proposition that was previously underivable and had thus been classified as false. Further areas of application include reasoning about

natural kinds, diagnostic reasoning, and natural language processing (Reiter, 1987; McCarthy, 1986).

Work on formalizing nonmonotonic inference has progressed rapidly since its beginnings in the 1970s, and there is now a large body of mature work in the area, much of it highly technical in character. (For a collection of seminal papers, see Ginsberg, 1987; for surveys of the field, see Brewka *et al.*, 1997, and the articles in Gabbay *et al.*, 1994. Antoniou, 1997, offers a relatively accessible introduction to the area.)

One of the most important nonmonotonic formalisms is *default logic*, developed by Raymond Reiter (1980). This involves supplementing first-order logic with new rules of inference called *default rules*, which have the form in (1).

$$(1) \frac{p : q}{r}$$

P is known as the *prerequisite*, q as the *justification*, and r as the *consequent*. Such rules are to be read, 'If p , and if it is consistent with the rest of what is known to assume that q , then conclude that r .' In simpler cases ('normal defaults'), q and r are the same, so the rule says that given the prerequisite, the consequent can be inferred, provided it is consistent with the rest of one's data. Thus the rule that birds typically fly would be represented as (2).

$$(2) \frac{Bird(x) : Flies(x)}{Flies(x)}$$

This says that if x is a bird and the claim that x flies is consistent with what we know, then we can infer that x flies. Given that all we know about Tweety is that he is a bird, we can therefore infer that he flies. The inference is nonmonotonic, since if we subsequently acquire information that is inconsistent with the claim that Tweety flies, then the rule will cease to apply to him.

The application of default rules is tricky, since it is necessary to check their justifications for consistency not only with one's initial data but also with the consequents of any other default rules that may be applied. The application of one rule may thus block that of another. To solve this problem, Reiter introduced the notion of an *extension* for a *default theory*. A default theory consists of a set of premises W and a set of default rules D . An extension for a default theory is a set of sentences E that can be derived from W by applying as many of the rules in D as possible (together with the rules of deductive inference) without generating inconsistency. An extension of a default theory can be thought of as a reasonable development of it.

Another approach, one closely related to default logic, is *autoepistemic logic* (Moore, 1985). This turns

on the idea that we can infer things about the world from our introspective knowledge of our own minds (hence the term 'autoepistemic'). From the fact that I do not believe that I owe you a million pounds, I can infer that I do not owe you a million pounds, since I would surely know if I did. Building on this idea, autoepistemic logic represents rules of thumb as implications of claims about one's own ignorance. For example, the rule that birds typically fly can be represented as the conditional claim that if something is a bird and one does not believe that it *cannot* fly, then it does fly. Given introspective abilities, one can use this claim to draw the defeasible conclusion that Tweety flies, based on one's ignorance of reasons to think he cannot. This approach can be formalized by using the apparatus of modal logic, with the modal operator L interpreted as 'It is believed that.'

A third approach is *circumscription* (McCarthy, 1980, 1986; see also Lifschitz, 1994). This involves formulating rules of thumb with abnormality predicates and then restricting the extension of these predicates – *circumscribing* them – so that they apply to only those things to which they must apply, given the information currently available. Take the Tweety case again. We render the rule of thumb that birds typically fly as the conditional in (3), where 'Abnormal' signifies abnormality with respect to flying ability.

$$(3) \forall x (Bird(x) \ \& \ \neg Abnormal(x) \rightarrow Flies(x)).$$

This does not, of course, allow us to infer that Tweety flies, since we do not know that he is not abnormal with respect to flying ability. But if we add axioms that circumscribe the abnormality predicate so that it applies to only those things that are currently known to be abnormal in this way, then the inference can be drawn. This inference is nonmonotonic, since if we were to add the premise that Tweety *is* abnormal with respect to flying ability, then the extension of the circumscribed predicate would expand to include Tweety, and the inference would no longer go through. Unlike the other strategies mentioned, circumscription can be formulated by using only the resources of first-order predicate calculus, though its full development requires the use of second-order logic, allowing quantification over predicates.

Each of these approaches has its own strengths and weaknesses, but there are some general issues that affect them all. One problem is that all of them allow for the derivation of multiple incompatible sets of conclusions from the same premises. In default logic, for example, a theory may have different extensions depending on the order in

which the rules are applied. The standard example is the theory that consists of the premises *Nixon is a Quaker* and *Nixon is a Republican*, together with the default rules *Quakers are typically pacifists* and *Republicans are typically not pacifists*. Each of these rules blocks the application of the other, since the consequent of one is incompatible with the justification of the other. The theory thus has two incompatible extensions – one in which the first rule is applied and containing the conclusion that Nixon is a pacifist, the other in which the second rule is applied and containing the conclusion that Nixon is not a pacifist. Similar results occur with the other approaches mentioned.

The existence of multiple extensions is not in itself a weakness – indeed, it can be seen as a strength. We might regard each extension as a reasonable extrapolation from the premises and simply plump for one of them. (This is known as the *credulous* strategy. The alternative *skeptical* strategy is to endorse only those claims that appear in every extension.) In some cases, however – particularly in reasoning involving time and causality – a plausible theory generates an unacceptable extension. Much work has been devoted to this problem – one strategy being to set priorities among default rules, which determine the order in which they are applied and so restrict the conclusions that can be drawn. (A much-discussed case is the *Yale shooting problem*, introduced in Hanks and McDermott, 1987, where the rule of thumb that living things typically stay alive generates an unexpected conclusion in reasoning about the effects of using a firearm. For a description of the scenario, see **Frame Problem**, and for discussion, see Shanahan, 1997.)

A second problem concerns implementation. The goal of much work in this area is to build artificial nonmonotonic reasoning systems, but the formal approaches that have been devised are not easy to implement. Default logic, for example, requires checking sets of sentences for consistency, and there is no general procedure for computing such checks. Restricted applications have been devised, however, and aspects of nonmonotonic reasoning have been effectively implemented by using the techniques of *logic programming* (programming in languages based on formal logic).

A third issue concerns the piecemeal character of much work in this area. Theories have been developed and elaborated in response to particular problem cases and with relatively little attention to the features of the inference relations they generate. Recent work has begun to address this issue and to identify properties that are desirable in any nonmonotonic inference relation (see, for example, Makinson, 1994). One of

these is that adding a conclusion back into the premise set should not undermine any other conclusions – a property known as *cautious monotonicity*.

Finally, a word about probabilistic logics. Most nonmonotonic formalisms embody a *qualitative* approach: premises and conclusions are treated as either true or false, as in deductive logic. But probabilistic logics, in which propositions are assigned continuous probability values, can also be used to model certain types of nonmonotonic inference. In probabilistic reasoning, it is common to treat a conclusion as warranted if its probability, given the premises, exceeds a certain threshold. This inference relation is nonmonotonic, since the addition of new premises may lead to a readjustment of one's probability assignments, with the result that a conclusion that previously passed the threshold now falls short of it. However, although probabilistic logics are well suited for modeling reasoning under uncertainty (see Shafer and Pearl, 1990), it is unlikely that they can do all the work required of a theory of nonmonotonic inference (McCarthy, 1986).

See also: Conditionals; Frame Problem; Logic and Language: Philosophical Aspects; Logical Consequence; Monotonicity and Generalized Quantifiers.

Bibliography

- Antoniou G (1997). *Nonmonotonic reasoning*. Cambridge, MA: MIT Press.
- Brewka G, Dix J & Konolige K (1997). *Nonmonotonic reasoning: an overview*. Stanford: CSLI Publications.
- Gabbay D M, Hogger C J & Robinson J A (eds.) (1994). *Handbook of logic in artificial intelligence and logic programming, vol. 3: Nonmonotonic reasoning and uncertain reasoning*. Oxford: Oxford University Press.
- Ginsberg M L (ed.) (1987). *Readings in nonmonotonic reasoning*. Los Altos, CA: Morgan Kaufmann.
- Hanks S & McDermott D (1987). 'Nonmonotonic logic and temporal projection.' *Artificial Intelligence* 33, 379–412.
- Lifschitz V (1994). 'Circumscription.' In Gabbay, Hogger & Robinson (eds.), 298–352.
- Makinson D (1994). 'General patterns in nonmonotonic reasoning.' In Gabbay, Hogger & Robinson (eds.), 35–110.
- McCarthy J (1980). 'Circumscription – a form of nonmonotonic reasoning.' *Artificial Intelligence* 13, 27–39. [Reprinted in Ginsberg (1987).]
- McCarthy J (1986). 'Applications of circumscription to formalizing common-sense knowledge.' *Artificial Intelligence* 28, 89–116. [Reprinted in Ginsberg (1987).]
- McCarthy J & Hayes P J (1969). 'Some philosophical problems from the standpoint of artificial intelligence.' In Meltzer B & Mitchie D (eds.) *Machine intelligence*,

- vol. 4. Edinburgh: Edinburgh University Press. 463–502. [Reprinted in Ginsberg (1987).]
- Moore R C (1985). ‘Semantical considerations on nonmonotonic logic.’ *Artificial Intelligence* 25, 75–94. [Reprinted in Ginsberg (1987).]
- Reiter R (1978). ‘On closed world data bases.’ In Gallaire H & Minker J (eds.) *Logic and data bases*. New York: Plenum. 55–76. [Reprinted in Ginsberg (1987).]

- Reiter R (1980). ‘A logic for default reasoning.’ *Artificial Intelligence* 13, 81–132. [Reprinted in Ginsberg (1987).]
- Reiter R (1987). ‘Nonmonotonic reasoning.’ *Annual Review of Computer Science* 2, 147–186.
- Shafer G & Pearl J (1990). *Readings in uncertain reasoning*. San Mateo, CA: Morgan Kaufmann.
- Shanahan M (1997). *Solving the frame problem*. Cambridge, MA: MIT Press.

Nonstandard Language Use

A Bezuidenhout, University of South Carolina, Columbia, SC, USA

© 2006 Elsevier Ltd. All rights reserved.

To talk of *non-standard* language use presupposes that we understand what it is for a use of language to be *standard*. A standard use can be thought of either as a use that conforms to some standard (viz. to a linguistic rule or convention) or as a use that is the usual or most common one. So, a non-standard use is either one that flouts a linguistic convention or that is an uncommon or novel use. Various linguistic phenomena somehow marked as non-standard will be discussed and classified as belonging to one or another of these two categories of non-standard usage. A possible third sense of ‘non-standard,’ which is the opposite of Bach’s (1995) notion of a *standardized* use, also will be discussed. Although non-standardized uses may be *pragmatically marked* uses of expressions, they don’t flout linguistic conventions and are not novel. Speech errors, such as cases in which a speaker says ‘pig vat’ instead of ‘big fat,’ also could be viewed as cases of non-standard language use, but will not be discussed in this article.

A conventional use, one conforming to a linguistic convention, is an agreed-upon use. The agreement needn’t be an explicit one but simply a matter of members of a linguistic community conforming their use to that of other members of the community, so long as others do the same. Malapropisms are one sort of example of *non-conventional* use. Malapropisms are cases where a speaker (unintentionally) substitutes a word for another word, whose agreed-upon meaning is different from what the speaker intends to convey. Usually the substitution is based on some sort of sound similarity between the ‘correct’ and ‘incorrect’ usage.

In his play *The Rivals*, Richard Sheridan has a lot of fun at the expense of his character Mrs. Malaprop. For example, Mrs. Malaprop exclaims to Sir Anthony Absolute, ‘you surely speak laconically!’, meaning to

surmise that Sir Anthony is speaking ironically. A little later she hopes that Sir Anthony will regard her niece, Lydia Languish, ‘as an object not altogether illegible,’ meaning that she hopes her niece will not be regarded as someone ineligible for a match with Sir Anthony’s son.

Such malapropisms are not confined to fiction. For example, in some experimental work on referential communication reported by Brown (1995), a participant in a map-reading task referred to an electric pylon as a colon. Another example occurred on the June 16, 2004, broadcast of the Tavis Smiley Show on National Public Radio. One of Smiley’s guests praised him by saying, ‘You have spoken up so unanimously for black people.’ The concept of unanimity only makes sense when one is talking about a group of people, and Smiley is a single individual. However, what the guest was thinking is that whenever Smiley speaks up he speaks up for black people. So, if one thinks of Smiley and his past selves as a group, one can think of them as speaking with one voice and hence as unanimous. This last example shows that malapropisms are not always based on *sound* similarities but can arise from quite complicated cognitive comparisons (e.g., Smiley and his past selves are like a group of individuals).

Malapropisms are extreme examples of misuse, and rather idiosyncratic. There are other examples of ‘misuse’ that are more widespread among members of a linguistic community. For example, many people use the sentence, ‘Hopefully, Bush will be defeated’ to express their hope that Bush will be defeated, not the thought that the defeating of Bush will be performed with hope. Or again, many people say things such as, ‘Me and him are going to the movies’ or ‘I could care less’ or ‘Everyone will be bringing their spouses’ instead of ‘He and I are going to the movies’ or ‘I couldn’t care less’ or ‘Everyone will be bringing his or her spouse.’

Given how widespread these latter ‘misuses’ are, one might argue that, rather than being misuses of English, they are acceptable uses within some

'non-standard' dialects of English. Calling these dialectal varieties 'non-standard' is rather contentious. There are some language purists who argue that only 'standard' English is acceptable usage. However, as Chomsky (2000) points out, calling some varieties of English 'standard' and regarding them as normative for English usage generally has nothing to do with the linguistic properties of these language varieties, but is purely a matter of power politics. The dialect of the ruling group is inevitably regarded as the norm for correct speech. Everyone has heard the old saw about a language being a dialect with an army and a navy. But as Chomsky remarks: "To say that one variety of English is 'right' and another 'wrong' makes as much sense as saying that Spanish is right and English is wrong" (2000: 71).

According to the second sense of 'non-standard' articulated above, non-standard uses are novel or 'nonce,' i.e., one-off, uses. The largest class of uses that have been thought of as 'nonce' are the uses of sentences to convey Gricean *particularized* conversational implicatures. For instance, a mother asks her teenage son to tidy his room and he replies: 'I have to beat this boss real quick.' The son is explicitly talking about what he has to do in the video game he's playing but is implicitly refusing to clean his room right then. In this sense of 'nonce' meaning, the novel meaning is something that is merely *indirectly* or *implicitly* conveyed. Such nonce meanings are conveyed because the hearer understands the speaker to have explicitly said one thing but to have meant something else (either in addition or instead). Thus, implicatures are not nonce meanings *for expressions*. Rather, they are novel meanings conveyed *by* the use of expressions that are understood to have their conventional meanings.

Metaphorical, ironic, metonymic, and other such 'non-literal' uses of expressions have also been thought of as nonce uses. For example, 'The soldiers butchered the villagers' can be used metaphorically to express the fact that the soldiers killed the villagers in a particularly inhumane way. 'You're a fine friend' can be used ironically to express the fact that the hearer has done something unfriendly. 'The ham sandwich is getting restless' can be used metonymically to express the fact that the orderer of the ham sandwich is getting restless. Many philosophers analyze such uses in terms of Gricean particularized conversational implicatures. The speaker says one thing but indirectly conveys something else. Inasmuch as such uses are analyzed in Gricean terms, these again would not be cases in which *expressions* are given novel meanings. Rather, *by* using expressions with their conventional meanings, speakers *indirectly* express novel meanings.

(Of course, the examples of metaphor, irony, and metonymy given above are very hackneyed, and in

this sense are not 'novel.' For example, the 'butcher' metaphor has become a 'dead' metaphor. Many dictionaries of English usage now list 'to kill in a brutal manner' as one of the meanings of 'to butcher.' Other, 'fresher,' examples could have been given, but these would have required more context setting.)

Not everyone would agree that cases of metaphor and metonymy are correctly analyzed in Gricean terms. An alternative 'contextualist' view denies that we interpret the entire sentence literally and only then (because of a violation of some Gricean maxim, such as the maxim of quality) derive an implicated metaphorical meaning. For example, consider the following

- (1) Dickens is easy to read.
- (2) The appendectomy is on top of the cabinet.
- (3) The stinkbug has left the room.

In (1), which is an example of a producer-for-product metonymy, 'Dickens' is used to refer to the novels that Dickens wrote. In (2), one can imagine a context in which 'the appendectomy' is used to refer to the file of the patient who is scheduled for an appendectomy. And in (3), 'the stinkbug' can be used metaphorically to refer to a person whose personal hygiene leaves much to be desired. According to the contextualist, local pragmatic processes operate on the referring terms ('Dickens,' 'the appendectomy,' 'the stinkbug') to yield pragmatically determined contents. These become a part of the propositional contents that are directly expressed by sentences (1)–(3) in their conversational contexts. For example, the speaker of (2) doesn't say falsely that a medical procedure is on the cabinet and thereby indirectly convey the proposition that the file of a patient scheduled for an appendectomy is on the cabinet. Rather, the latter proposition is directly expressed. See Carston (1997) and Bezuidenhout (2001) for a defense of the contextualist view of metaphor.

If this view is correct, there is a sense in which the expressions 'Dickens,' 'the appendectomy,' etc. are given novel meanings in different contexts. These novel meanings are not merely indirectly conveyed *by* literally saying something else. Note, this view is *not* committed to a 'Humpty Dumpty' theory of meaning, according to which expressions can mean whatever their users intend them to mean. (In *Through the Looking-Glass*, Humpty Dumpty says to Alice, 'There's glory for you' and claims to mean 'There's a nice knockdown argument for you.') According to the contextualist view, the novel contextual meanings of expressions are pragmatic developments of semantically encoded meanings. The local pragmatic processes that determine contextual meanings are thus not completely unconstrained.

Speakers cannot convey novel meanings that are completely unrelated to conventionally encoded meanings. (For a more radical view, see Donnellan [1968] and Davidson [1986]. They argue that, in appropriate contexts, 'glory' *can* mean *a nice knock-down argument*.)

Another class of uses that have been called 'nonce' are ones in which speakers coin new words. Attested examples are as follows, the first two given by Clark (1992: 315–316), the third by Higginbotham (2002: 577):

- (4) He's being grand juried for possible conflict in real estate loans.
- (5) Today I'm going gallerying.
- (6) I've hard-driven the diskette.

These novel uses are similar to ones that have already entered common usage, such as 'He docked the boat' or 'The librarian shelved the books.' These uses all obey a productive rule, in which a noun is turned into a verb by the addition of a verbal particle ('-ed' to signal past tense, '-ing' to signal an ongoing or incomplete action). The novel words are coined because there is no readily available, short expression already in common use that quite covers the speaker's intended meaning.

In some cases, people invent new words when there are already perfectly good words in common usage that mean exactly what the speaker intends. For example, on the June 16, 2004, edition of the *Tavis Smiley Show* on NPR, one of Smiley's guests said: 'You have to converse and talk to your child on a man to man level.' Presumably what was meant was that one should *converse* with one's child. What is already a verb, 'to converse,' is subjected to a morphological rule for verb formation, by the addition of the suffix '-ate,' perhaps on analogy with the verb 'to orientate,' which is frequently used instead of the plain 'to orient' (this latter being the norm in standard American English, although not in standard British English, which favors 'to orientate').

A third sense of 'standard use' should be mentioned. It corresponds to what Bach (1995) calls a *standardized* use, which he distinguishes from a conventional use. One reason that Bach appeals to his notion of standardized use is to counter attempts by relevance theorists such as Carston (1995) to reconstrue Gricean *generalized* conversational implicatures as part of what is directly expressed. Consider, for example, the Gricean treatment of scalar terms, such as numerals. On this view, 'Grice has three teeth' literally says that Grice has *at least* three teeth, but it conversationally implicates in a generalized way that Grice has *at most* three teeth. These two propositions

together entail that Grice has *exactly* three teeth. Carston's alternative view is that 'three teeth' is semantically underspecified. When used in an appropriate context, it can be specified in such a way that what is said (i.e., what is explicitly conveyed) in that context by 'Grice has three teeth' is that *Grice has exactly three teeth*.

Bach offers a third alternative. On this view, 'three teeth' has a standardized use to convey *exactly three teeth*. A standardized use is something like a default use. It is a default in the sense that it has become a regular practice – perhaps a mental habit – for members of the linguistic community to use the word(s) in question in a certain way, given that circumstances are normal. Only if circumstances are somehow unusual will a different interpretation be called for. Bach denies that the standardized meaning is explicitly conveyed. It is implicitly conveyed (he calls it an 'implicature'), but because it is conveyed by default in normal circumstances, it gives the illusion of being directly conveyed. The inference that is needed to derive this standardized interpretation has become compressed by precedent. Hence, hearers are not aware of having made any inference. A similar view, according to which generalized implicatures are default inferences, has been defended by Levinson (2000).

The details of Bach's views are not important here. The main point is to introduce a third sense of 'non-standard use.' It corresponds to a *non-standardized* use in Bach's sense. A non-standardized use is one that is called for when circumstances are somehow not 'normal' and hence the default meaning must be overridden. For instance, suppose Grice has a large collection of shark teeth and you need exactly three shark teeth for an artwork you are creating. Believing that Grice might be willing to give up some of his shark teeth for you I say 'Grice has three teeth.' Arguably what is conveyed here is that Grice has at least three teeth, not that he has exactly three. Now, according to Griceans such as Bach and Levinson, the conventionally encoded meaning of 'three teeth' is *at least three teeth*. In other words, when circumstances are *not* normal and the default meaning is overridden, the meaning that is conveyed is the literally encoded meaning. So the non-standardized use in this case is the conventional one (i.e., it is a standard use in the first sense articulated above).

In the example given in the previous paragraph, pragmatic context indicated that a non-default understanding of 'three teeth' was called for. Sometimes, however, the linguistic expression itself signals whether it is to be given a default or non-default reading. There are pragmatically marked and unmarked ways of saying things. Unmarked ways of

saying things call for default interpretations. Marked ways of saying things call for non-default interpretations. For example, there is a difference between saying 'The policeman stopped the car' and 'The policeman brought the car to a stop.' The former suggests that the stopping was the normal sort of stopping, whereas the latter suggests that the policeman did something unusual to halt the car. But although non-standardized uses may be *pragmatically marked* uses of expressions, they don't flout linguistic conventions and they are not novel uses, or at least no more novel than their unmarked alternatives. Both 'The policeman stopped the car' and 'The policeman brought the car to a stop' will call for quite a lot of pragmatic inferencing to understand what the speaker said.

See also: Conventions in Language; Metaphor: Philosophical Theories; Pragmatic Determinants of What Is Said.

Bibliography

Bach K (1995). 'Standardization vs. conventionalization.' *Linguistics and Philosophy* 18, 677–686.

- Bezuidenhout A (2001). 'Metaphor and what is said: a defence of a direct expression view of metaphor.' *Midwest Studies in Philosophy* 25, 156–186.
- Brown G (1995). *Speakers, listeners and communication: explorations in discourse analysis*. Cambridge: Cambridge University Press.
- Carston R (1995). 'Quantity maxims and generalized implicature.' *Lingua* 96, 213–244.
- Carston R (1997). 'Enrichment and loosening: complementary processes in deriving the proposition expressed?' *Linguistische Berichte* 8, 103–127.
- Chomsky N (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Clark H (1992). *Arenas of language use*. Chicago: University of Chicago Press.
- Davidson D (1986). 'A nice derangement of epitaphs.' In Lepore E (ed.) *Truth and interpretation: perspectives on the philosophy of Donald Davidson*. Oxford: Blackwell. 433–446.
- Donnellan K (1968). 'Putting Humpty Dumpty together again.' *Philosophical Review* 77, 203–215.
- Higginbotham J (2002). 'On linguistics in philosophy and philosophy in linguistics.' *Linguistics and Philosophy* 25, 573–584.
- Levinson S (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Normativity

S Gross, University of Pennsylvania, Philadelphia, PA, USA

© 2006 Elsevier Ltd. All rights reserved.

Normative claims concern the ways things should be. Nonnormative (or descriptive) claims concern how things are, which can diverge from how they should be. 'No one should speed' is a normative claim; 'they are speeding' is not. Neither is '85% of drivers speed,' even if it describes how things in fact normally are. Statistical norms must be distinguished from norms or standards for how things ought to be.

In exploring the relevance of normativity to language, it is useful to bear in mind that normative claims fall into various kinds and can possess a range of further features. Consider the following examples:

1. Sheets should be clean.
2. Friends should not spread gossip.
3. Smoking is permitted.
4. If you believe some claim, you shouldn't believe its negation.
5. Only goalkeepers may touch the ball with their hands.

6. Goalkeepers should avoid straying from their goal.
7. Pedestrians may cross only at the crosswalk.
8. You shouldn't wear blue slacks with a brown jacket.
9. If you want to improve, you should practice more.

These examples illustrate a variety of features of normativity. Normative claims may state, not only how things should be (1), but also how they should *not* be (2), as well as how they are allowed though not necessarily required to be (3). The subject of norms can include agents (2), actions (3), and states of affairs (1). Normative claims can be conditional (4). The applicability of a norm may be constitutive of an activity ((5) – one is not playing soccer if players may use their hands). Or it may just be constitutive of doing it well (6). Normative claims are not restricted to moral claims; they can be legal (7), aesthetic (8); means–end (9); etc.

Issues involving normativity arise regarding language and linguistics in a variety of ways. Four particularly prominent questions are discussed below: May linguistics itself issue normative pronouncements concerning language and language use? Are semantic

properties inherently normative? Are speech acts and pragmatic phenomena more generally inherently normative? What linguistic properties do the terms used to express normative claims themselves have?

Prescriptive Versus Descriptive Linguistics

Language users often turn to dictionaries to answer questions concerning how a word should be used. They view dictionaries as compendia of prescriptions, not descriptions. Linguists, however, generally do not aspire to normative conclusions in their research. They aim to describe and explain linguistic phenomena. On their view, linguistics, at least as they practice it, is and ought to be an empirical science. To the normative question ‘what sort of pursuit should linguistics be?’ they thus provide an answer that eschews normativity so far as the pronouncements of linguistics are concerned (see **Description and Prescription**).

It is a distinct question, however, whether the scientific study of language, though it should not enter normative claims, must also prescind from examining and adverting to normative phenomena. Canons of politeness, for example, are normative. But they impact language and language use. For instance, they affect turn-taking in discourse and can become lexically, morphologically, and syntactically encoded. They thus seem important objects of study for linguists. Some philosophers of science suggest that naturalistic inquiry cannot accommodate intrinsically normative phenomena. Others object that this represents an overly narrow conception of naturalistic inquiry. This is an instance of a larger (venerable) debate concerning the status of the ‘human’ (or, social) sciences and their relation to the natural sciences (cf. Martin and McIntyre, 1994: Part III).

Semantics and Normativity

It is often claimed that semantic properties are intrinsically normative, in some manner encoding standards of correctness. Though fleshed out in various ways, the defense typically involves arguments purporting to show that the possession of a semantic property entails or is entailed by something normative.

For example, it is sometimes suggested that, if ‘dog’ denotes dogs, then one ought to apply the term only to dogs. If one applies the term to a cat, one has made a mistake (cf. Kripke, 1982). Here, it is claimed that possession of a semantic property has normative consequences. However, it is not generally the case that,

if one should or should not do something with an X, then Xs are intrinsically normative; one ought not throw rocks at people, but rocks are not intrinsically normative – nor is the property of being a rock or any of a rock’s intrinsic properties. What matters is whether the normative conclusion follows *just* from the thing’s being X. That one should not throw rocks at people follows, not just from what rocks are like, but from that and one’s obligation not to cause bodily harm.

The question is thus whether the possession of a semantic property like ‘denotes dogs’ is sufficient in and of itself to generate obligations. Arguably, it is not: that one ought to apply ‘dog’ only to dogs would seem to follow, not just from its possessing the semantic property of denoting dogs, but from that and the further assumption that one ought to aim at truth. Perhaps one indeed ought to aim at truth; perhaps not always (maybe there are circumstances in which lying is justified – for example, to confuse a would-be cat-killer). But if the assumption that one should aim at truth must be added, then the normative conclusion does not follow from the semantic claim alone (Horwich, 1998: Chapter 8).

Others claim that semantic properties are intrinsically normative because *what makes it the case* that terms have their semantic properties are certain normative phenomena. The possession of semantic properties is thus alleged to be entailed or determined by the obtaining of normative facts. For example, some argue that semantic claims (such as that ‘dog’ denotes dogs) obtain in virtue of a language user’s being disposed to apply ‘dog’ to dogs in appropriate circumstances – where what makes circumstances appropriate may include how things optimally ought to be when applying such a term. Others, however, attempt to show that, contrary to such claims, one can ‘naturalize’ semantic content. They maintain that terms have their semantic properties in virtue of facts that can be characterized nonnormatively (cf. Loewer, 1996).

One project in foundational semantics that would reject ‘naturalization’ thus construed is inferentialism (Brandom, 1994). According to it, semantic claims (such as ‘dog’ denotes dogs) obtain in virtue of it being the case that certain inferences involving the term would be correct to draw, and certain inferences would not. Logical terms provide the best cases for this approach: the claim would be that ‘and’ means what it does in virtue of the validity of inferences of the form (1) ‘A and B’ implies A and (2) A, B implies ‘A and B.’ Among the challenges is to extend this strategy to other terms. In the case of ‘dog,’ the relevant inferences would include not only transitions (in appropriate circumstances) from the presence of dogs

to ‘Those are dogs,’ but also the inference from ‘Those are dogs’ to ‘Those are mammals’ – and much more besides.

Pragmatics and Normativity

Even if the meanings of expressions are not inherently normative, one might claim that what speakers do with expressions is inherently normative. Human language use, at least in core cases, is a species of intentional action, and all such action is done for reasons. One’s reasons for acting are subject to normative assessment; they may or may not be *good* reasons. The intentional nature of linguistic action figures prominently in speech act theory (Austin, 1975) and accounts of conversational implicature (Grice, 1989). It also provides one motivation for developing game-theoretic models of language use.

Consider first the study of speech acts. It is often claimed that, for the utterance of a sentence to constitute the performance of a speech act, it must be performed intentionally by a speaker beholden to certain norms, with specific norms attaching to specific speech acts. For example, for the utterance of a particular sentence to constitute an assertion, the speaker must represent herself as having warrant for the truth of the claim asserted, with the result that she can be held accountable if the claim is not true or at least was not asserted on sufficient grounds. The transmission of truth, or perhaps knowledge, is said to be the aim or point of the practice of assertion. In this case, the norms constraining linguistic action would be constitutive of the kind of action at issue (cf. Williamson, 1996).

In performing a specific speech act such as assertion, it is often one’s intention to communicate more than just what one asserts. (Sometimes one intends to communicate something *instead* of what one makes as if to assert.) For example, in answer to the question ‘Do you think John is coming?’ one might utter the sentence ‘There’s a lot of snow on the roads’ in order both to assert that there’s a lot of snow on the roads and to communicate that John is probably not coming. Grice (1989) argued that such ‘conversational implicatures’ are possible because of language users’ sensitivity to the reasons for which a cooperative speaker would utter particular sentences in particular conversational contexts. He articulated a set of maxims for cooperative language use – for example, that one be as informative as possible, all else being equal, but also that one refrain from prolixity, all else being equal – that parties to a conversation tacitly assume one another to be observing. Speakers can then rely on hearers to infer an implicature, as in the case above, when doing so preserves the assumption of

cooperation. Such maxims are norms that specifically apply to cooperative conversations and that enable speakers and hearers to discern the specific conversational reasons behind linguistic actions.

In the preceding discussion, statistical norms were sharply distinguished from norms for how things ought to be. But statistical norms, like other nonnormative facts, can have normative upshot; how things are affects what means one should take to achieve one’s ends. Conforming to a trend, for example, can in some cases increase the likelihood of obtaining an outcome. This is certainly so with language. Suppose one has an interest in one’s utterances being understood. One then has reason to speak in a way that will promote understanding. But then, insofar as there exist statistical norms concerning pronunciation, assignment of meaning, expected prolixity, and other linguistic matters, there is reason to conform to those norms, since those norms will correlate with other language users’ expectations. This provides one rationale for using game-theoretic techniques to study both the development of linguistic norms (in the statistical sense) and the ways speakers deploy language on particular occasions (see, e.g., Nowak and Komarova 2001, and with reference to pragmatic phenomena, Parikh, 2001).

Linguistic Properties of Normative Terms

We express normative claims – like any other claims – by using language. Normative expressions such as ‘ought,’ ‘justified,’ and ‘you should not do that’ are thus themselves proper objects of linguistic theorizing. Three examples are the use of deontic logic to capture the content of normative lexical items, expressivism as a non-truth-conditional analysis of normative sentences, and the study of generics.

Deontic logic – a branch of intensional logic – studies the logical relationships among normative expressions. ‘It’s obligatory that one X,’ for example, implies ‘It’s not forbidden that one X.’ Logicians studying such expressions construct formal languages with axioms and inference rules that capture these relations. Such investigations can be construed as contributions to the lexical semantics of the normative terms (see Føllesdal and Hilpinen, 1971).

Expressivism in its most basic form is the doctrine that normative claims such as ‘Murder is wrong’ do not purport to determine truth-conditions, but rather express an attitude toward a certain kind of action – in this case a negative attitude toward murder. Normative claims on this view are thus no more liable to assessment for truth or falsity than such aesthetic expressions as ‘Chocolate: yum!’ Expressivism might

be motivated by the intuition that normative claims – for example, moral claims – do not purport to express objective facts. Expressivism, however, has difficulty accommodating embedded normative claims. The sentence ‘If murder is wrong, then what Mary did was wrong’ does not itself express any attitude to murder. This would seem to indicate that the expression of an attitude is not itself among the semantic properties of the expressions used. Moreover, the following argument seems deductively valid: murder is wrong; if murder is wrong, then what Mary did was wrong; therefore, what Mary did was wrong. It’s natural to elucidate this validity by viewing the argument as an instance of the truth-preserving inference schema *Modus Ponens*. But this can seem blocked if normative claims do not determine truth-conditions. (For discussion from the perspective of a more sophisticated version of expressivism, see Gibbard, 2003.)

Generic claims – such as ‘Dogs have four legs’ – provide an example of unobvious normativity. The sample sentence’s truth, on the relevant reading, does not depend on the four-leggedness of *all* dogs. (Amputees do not witness its falsity.) Nor does the existence of *some* four-legged dogs suffice for its truth. The claim is not even that *most* dogs have four legs: a generic claim does not report a statistical norm. (It may be true that a spider’s life has four stages, even if the vast majority of spiders never make it past the first.) The content of a generic claim is rather normative; *normal* dogs have four legs. Perhaps this is even in some sense how dogs ought to be – at least to conform to conversationally relevant expectations. (On generics generally, see Carlson and Pelletier, 1995.)

See also: Assertion; Causal Theories of Reference and Meaning; Description and Prescription; Generic Reference; Linguistics as a Science; Reference: Philosophical Theories.

Bibliography

- Austin J L (1975). *How to do things with words* (2nd edn.). Cambridge, MA: Harvard University Press.
- Brandom R (1994). *Making it explicit*. Cambridge: Harvard University Press.
- Carlson G & Pelletier F J (eds.) (1995). *The generic book*. Chicago: University of Chicago Press.
- Føllesdal D & Hilpinen R (1971). ‘Deontic logic: an introduction.’ In Hilpinen R (ed.) *Deontic logic: introductory and systematic readings*. Dordrecht: Reidel. 1–35.
- Gibbard A (2003). *Thinking how to live*. Cambridge, MA: Harvard University Press.
- Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Horwich P (1998). *Meaning*. Oxford: Oxford University Press.
- Kripke S (1982). *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.
- Loewer B (1996). ‘A guide to naturalizing semantics.’ In Hale B & Wright C (eds.) *The Blackwell companion to the philosophy of language*. Oxford: Blackwell. 108–126.
- Martin M & McIntyre L (eds.) (1994). *Readings in the philosophy of social science*. Cambridge, MA: MIT Press.
- Nowak M & Komarova N (2001). ‘Towards an evolutionary theory of language.’ *Trends in Cognitive Sciences* 5, 288–295.
- Parikh P (2001). *The use of language*. Palo Alto, CA: CSLI Publications.
- Williamson T (1996). ‘Knowing and asserting.’ *The Philosophical Review* 105, 489–523.

This page intentionally left blank



Object-Dependent Thoughts

S Crawford, Lancaster University, Lancaster, UK

© 2006 Elsevier Ltd. All rights reserved.

Some of our thoughts involve reference to particular individual entities. Philosophers call these kinds of thoughts ‘singular thoughts.’ *Russell was a great philosopher, you’re standing on my foot, I’m tired, that raccoon got into my garbage last night* – these are all singular thoughts because each involves reference to a particular thing: Russell, you, me, and a certain raccoon, respectively. As these examples indicate, singular thoughts are usually expressed by sentences containing proper names (e.g., ‘Russell’), indexicals (e.g., ‘you’ and ‘I’), and demonstrative expressions (e.g., ‘that raccoon’).

Singular Thoughts as Object Dependent

Some philosophers maintain that singular thoughts are *object dependent*, by which they mean that the intentional content of the thought essentially involves the object that it is about, in the sense that the thought content would not be available to a thinker were the object not to exist. More precisely, a singular thought is object dependent just in case its content is such that (1) its existence depends on the existence of the object thought about, and (2) its identity depends on the identity of the object thought about. For example, consider the thought *that raccoon got into my garbage last night*, had by me while spying a particular raccoon skulking in my backyard. According to the doctrine of object dependence, if, counterfactually, no raccoon had in fact been there to be singled out by me, owing perhaps to my delusional or hallucinatory state of mind – let us call this the “empty possibility” – then there would have been no singular thought content for me to entertain. Consequently, my psychological condition in this situation would be different from what it is in the actual situation. Moreover, if, counterfactually, my thought had singled out a qualitatively indistinguishable but numerically different raccoon instead – call this the “duplicate possibility” – then the resulting thought would have had a different

content from the content that my thought had in the actual situation. Again, my overall psychological state in this duplicate possibility is different from what it actually is. The implication here for linguistic meaning is that the meaning of sentences containing genuine singular terms (e.g., proper names, indexicals, and demonstrative expressions) depends on the singular terms in question successfully referring to objects. On this view, nonfictional sentences containing non-referring singular terms, such as empty or bearerless names, are meaningless, in the sense that they fail to express any thoughts.

The doctrine of object dependence is a species of the more general doctrine of *externalism* about thought content, according to which some states of mind are such that we can be in them only if we bear certain appropriate relations to other things in our environment, and thus is opposed to *internalism* about the mind, according to which the contents of our thoughts are never dependent on any relations between us and other things in our environment. (Some philosophers, such as Burge [1982], accept the general doctrine of externalism but reject object dependence.)

Epistemological Consequences of Object-Dependence

It is controversial which, if any, singular thoughts are object dependent. Arguably, first-person thoughts expressed with the indexical ‘I’ are object dependent: it seems obvious that if I did not exist, then the thought that I now express with the sentence *I’m tired* could not exist; moreover, no one else could have had the very same thought. But the thesis that singular thoughts expressed with proper names and demonstratives are object dependent has seemed paradoxical to some philosophers. For when the idea of object dependence is applied to these other types of singular thoughts, it runs up against a strongly held intuition about the nature of thought content, namely, that we have a kind of direct, noninferential knowledge of the contents of our thoughts, in the sense that we know, just by thinking, whether we are having a thought and, moreover, what thought we are having.

The doctrine of object dependence seems to contravene this intuition about the epistemology of thought.

For, first of all, condition (1) above allows the possibility that a thinker could suffer the illusion of entertaining a thought when he was not in fact doing so. If, unbeknown to me, I am in what we have been calling an empty possibility and am hallucinating a raccoon rather than actually seeing one, it may seem to me that I am having a singular thought, which I might try to express with the sentence *that raccoon got into my garbage last night*, even though I am not. But is this kind of cognitive illusion really possible? It is very tempting to think, against this, that if it seems to me as if I am having a thought with a certain content, then I am. Perhaps I might be mistaken about which object, if any, my thought is about – but how could I be mistaken about whether I was even thinking a thought at all?

Condition (2) has also seemed problematic. Consider what we have called the duplicate counterfactual possibility, in which I see a different raccoon, qualitatively indistinguishable from the one I actually see, and think *that raccoon got into my garbage last night*. In such a case, everything will seem the same to me: the duplicate raccoon does not appear to affect my conscious awareness in any way different from how the actual raccoon affects it. But is not subjective indistinguishability the criterion for sameness and difference of thought content? Opponents of object dependence argue that in order for there to be a genuinely psychological or mental difference between the two cases, this difference must impinge on my conscious awareness in some way. The object-dependent theorist denies this, arguing that it is the product of a mistaken internalist picture of the mind, a picture that the object-dependent theorist urges us to reject in favor of an externalist view. The debate between object-dependent theorists and their opponents is thus linked to a certain extent to the larger debate between internalism and externalism about thought content.

The Central Motivation for Object Dependence

A number of different considerations have been advanced in favor of an object-dependent conception of singular thought, and many involve a synthesis of key ideas of Frege and Russell (Evans, 1982; McDowell, 1977, 1984, 1986; McCulloch, 1989). Advocates of this form of object dependence are often labeled ‘neo-Fregeans,’ which can be confusing, because object-dependent singular thoughts are also often called ‘Russellian thoughts,’ so one needs to be aware of differing terminology here.

Perhaps the best way to appreciate the object-dependent theorists’ point of view is to begin by noting that they do countenance thoughts that in a certain sense concern particular individuals but that would be available to a thinker were those individuals not to exist. Moreover, there is a straightforward sense in which the contents of these kinds of thoughts would remain unaffected were duplicate objects substituted for the actual ones. Calling these kinds of thoughts object-independent thoughts, we can say that although they concern particulars, the relation between their contents and their objects is much less direct or intimate than the relation between the contents and objects of object-dependent thoughts (the idea goes back to Russell’s seminal distinction [1910–1911] between knowledge by description and knowledge by acquaintance).

The most obvious examples of object-independent thoughts are thoughts that involve definite description concepts, thoughts of the form *the F is G*. Consider the thought *the first man on the moon was an American*. As it happens, this thought is about Neil Armstrong because he was in fact the first man on the moon. But consider now the empty possibility in which the lunar landing was a hoax and the definite description *the first man on the moon* fails to designate anything. The object-dependent theorist holds that even though the thought fails to single out any actual object in the world, the thought still has a content, a content expressed, in part, by the definite description. Similarly, consider the duplicate possibility, in which Neil Armstrong’s identical twin is the first man on the moon. Despite the thought’s picking out a different man, the content of the thought remains the same – again, that expressed (in part) by the definite description *the first man on the moon*. The crucial point here is that the intentional content of the thought can be specified independently of the object, if any, that it is about.

The object-dependent theorist’s idea is this. Thought content is essentially representational: it represents the world as being a certain way; it lays down conditions that the world must meet in order for the thought to be true. That is to say, the content of a thought determines its truth conditions. In the case of a thought employing a definite description concept *the F* (a descriptive thought, for short), the thinker knows what those conditions are without knowing which object, if any, the thought concerns. If I say to you *the first man on the moon was American*, it is not necessary for you to know which object is the first man on the moon, nor even that there is such an object, in order for you to understand what I have said, in order for you to “grasp” the thought I expressed with this sentence. So long as

you understand all the words in the sentence and their mode of combination, you know exactly how the world is represented as being; you know what the thought is “saying” about reality. In other words, you know that the thought is true just in case there is a unique man who was first on the moon and who was American. It does not matter who this man happens to be – Neil Armstrong, his identical twin, or Buzz Aldrin. So long as there is such a man, the thought is true; and if there is no such man – either because no man at all has ever been on the moon or because more than one man stepped onto the moon at exactly the same time – then the thought is false. The representational content of a descriptive thought is thus independent of any object that the content might be about. The truth conditions make no reference to any man in particular.

When it comes to singular thoughts, however, the object-dependent theorist maintains that their representational content is not independent of any object the content is about. On the contrary, the content requires that a certain particular object be picked out. In order to understand or grasp the thought in question, one must know which particular object this is. Consider the foregoing example of a singular thought: *that raccoon got into my garbage last night*, based on my visual experience of a particular raccoon in my backyard (these kinds of singular thoughts are sometimes called ‘perceptual demonstrative thoughts’). Now, in having this thought, I am representing the world in a certain way. What way is this exactly? Well, I am not representing the world as merely containing a raccoon that got into my garbage last night, whichever raccoon that might be. The way I am representing the world as being involves that very raccoon. My thought is true just in case that raccoon (the very one I saw) got into my garbage last night; and in order for you to have this thought too, you need to know which particular raccoon is singled out by my perceptual demonstrative ‘that raccoon.’ Contrast this with the very different case in which I think the descriptive thought *the cleverest and boldest raccoon in the neighborhood got into my garbage last night*. All that it takes for this thought to be true is for there to be a unique raccoon, who is cleverer and bolder than all the rest and who got into my garbage – and you can grasp this thought without knowing which raccoon, if any, that was. If it turns out that there was no such raccoon, then my thought is straightforwardly false. But the truth conditions for my perceptual demonstrative thought make essential reference to the very object it is about. The truth or falsity of this thought of mine turns on the condition of a particular raccoon, namely, that raccoon – so that if there is no such creature, if (say) I am hallucinating,

there is nothing in the world to count as my thought being true or false. Consequently, in this empty possibility, my mental episode, whatever exactly its nature, has no truth conditions (for, to repeat, there is nothing *of which* I have judged to have a certain property; nor have I made the mere existential claim that there is *an* object with a certain property). Since thought content is essentially truth conditional, according to the object-dependent theorist, I have not in fact had a singular thought at all, only the illusion of one. Whether considerations like these in favor of object dependence apply equally to other kinds of singular thoughts, such as those expressed with proper names and indexicals (other than ‘I’), is a further question.

Criticisms and Rivals

Various criticisms have been leveled at the object-dependent conception of singular thought. Some of these arise from problems that the conception inherits from the general doctrine of externalism, such as its apparent conflict with certain features of self-knowledge (Davies, 1998). Three issues, however, stand out with respect to object dependence in particular.

The first is the question of what is going on, psychologically speaking, in the minds of deluded subjects in empty possibilities who suffer the illusion of entertaining singular thoughts. Their minds are not phenomenological blanks, after all; yet, according to the object-dependent theorist, they are not filled with any singular thoughts. Are such deluded subjects having any thoughts at all? If so, what kinds of thoughts are they having?

The second issue is closely related to the first and concerns the commonsense psychological explanation of the actions of deluded subjects. Normally, we explain agents’ actions – my charging into the backyard, say – by attributing singular thoughts to them – the belief that that raccoon got into my garbage last night, for example. But now consider my deluded duplicate who, after hallucinating a raccoon in the empty possibility, engages in the very same type of behavior of charging into the backyard. According to the object-dependent theorist, my duplicate here has no singular thought; that is, he has no belief the content of which is *that raccoon got into my garbage last night*. But, although he is hallucinating, his action is perfectly rational and so is presumably psychologically explicable by ordinary commonsense standards. But how do we so explain his behavior without attributing a singular thought to him (McDowell, 1977; Segal, 1989)? Moreover, if we can explain his behavior without attributing a singular thought to him, then why can we not do the same with me in the

actual situation? But if we can do this with me too, then it looks as if the ascription of object-dependent singular thoughts is 'psychologically redundant' – and that allegedly calls into question their very existence (Noonan, 1986, 1991; Segal, 1989).

The third issue, perhaps the most serious, is that there are powerful rival object-independent conceptions of singular thought, which are free of many of the problems that beset object-dependent theories. There tend to be two different kinds of alternative conceptions.

The first of these conceptions attempts to analyze singular thought content in wholly general or descriptive terms, in such a way that the same content can exist in duplicate and empty possibilities, in the manner of thoughts involving definite description concepts, discussed earlier (Schiffer, 1978; Searle, 1983, 1991; Blackburn, 1984: chapter 9). For example, we might try to analyze the content of the demonstrative expression *that raccoon* as equivalent to the content of the definite description *the raccoon I am seeing now* or *the raccoon causing this visual experience*.

The second approach opposes this kind of descriptive reduction and maintains a genuinely singular conception of singular thought but argues that a distinction between irreducibly singular (or "*de re*") content and object can still be drawn, again, in such a way that, as with the first alternative, the same singular content can exist in both duplicate and empty possibilities (Burge, 1977, 1982, 1983, 1991; Bach, 1987; Segal, 1989). This approach exploits an analogy between the semantics of sentences containing demonstratives and pronouns (*this is red*, *she is tall*) and the semantics of the open sentences of a logical system (*x is red*, *x is tall*) – namely, that both kinds of sentences are true or false only under an assignment of values to the demonstratives, pronouns, and free variables in question. The proposal is to treat a sentence such as *that is a raccoon* as like a predicate, or open sentence in the logician's sense, and to think of it as expressing a single content (a "propositionally incomplete" content) that is mentally applied, in different situations, to different objects, and even, in some situations, to no object at all.

These two alternatives each face their own difficulties, however. The first alternative seems to overintellectualize thinking. When I think *that raccoon got into my garbage last night*, I do not appear to be thinking about myself or the present moment or about causation or my own visual experiences, and even if I were doing so in a philosophical mood, it does not seem necessary for a creature to have such sophisticated concepts in order for it to have singular thoughts (McDowell, 1991; Burge, 1991; Searle, 1991). As for

the second alternative, it is not clear to what extent it departs from the intuitive principle that thought content is fully representational in the sense of always determining truth conditions. For in the empty counterfactual possibility, in which I hallucinate a raccoon, no value will be assigned to the demonstrative concept in my thought (*that raccoon*), and hence no truth conditions for the overall thought will be determined. The advocates of this second alternative approach thus seem committed to the view that I can have thoughts that possess no truth conditions, something that may give us pause.

See also: Counterfactuals; Descriptions, Definite and Indefinite: Philosophical Aspects; Dthat; Empty Names; Externalism about Content; Immunity to Error through Misidentification; Indexicality: Philosophical Aspects; Proper Names: Philosophical Aspects; Reference: Philosophical Theories; Truth Conditional Semantics and Meaning; Two-Dimensional Semantics.

Bibliography

- Bach K (1987). *Thought and reference*. Oxford: Clarendon Press.
- Beaney M (ed.) (1997). *The Frege reader*. Oxford: Basil Blackwell.
- Blackburn S (1984). *Spreading the word*. Oxford: Clarendon Press.
- Burge T (1977). 'Belief de re.' *Journal of Philosophy* 74, 338–362.
- Burge T (1982). 'Other bodies.' In Woodfield A (ed.) *Thought and object*. Oxford: Clarendon Press.
- Burge T (1983). 'Russell's problem and intentional identity.' In Tomberlin J (ed.) *Agent, language, and the structure of the world*. Indianapolis: Hackett.
- Burge T (1991). 'Vision and intentional content.' In Lepore E & Van Gulick R (eds.) *John Searle and his critics*. Oxford: Basil Blackwell.
- Carruthers P (1987). 'Russellian thoughts.' *Mind* 96, 18–35.
- Davies M (1998). 'Externalism, architecturalism, and epistemic warrant.' In Wright C, Smith B C & Macdonald C (eds.) *Knowing our own minds*. Oxford: Clarendon Press.
- Evans G (1982). *The varieties of reference*. McDowell J (ed.). Oxford: Clarendon Press.
- Frege G (1892). 'On Sinn and Bedeutung.' Black M (trans.). In Beaney (ed.).
- Frege G (1918). 'Thought.' Geach P & Stoothoff R H (trans.). In Beaney (ed.).
- McCulloch G (1989). *The game of the name*. Oxford: Clarendon Press.
- McDowell J (1977). 'On the sense and reference of a proper name.' *Mind* 86, 159–185.
- McDowell J (1984). 'De re senses.' *Philosophical Quarterly* 34, 283–294.

- McDowell J (1986). 'Singular thought and the extent of inner space.' In Pettit P & McDowell J (eds.) *Subject, thought and context*. Oxford: Clarendon Press.
- McDowell J (1991). 'Intentionality de re.' In Lepore & Van Gulick (eds.).
- Noonan H (1986). 'Russellian thoughts and methodological solipsism.' In Butterfield J (ed.) *Language, mind, and logic*. Cambridge: Cambridge University Press.
- Noonan H (1991). 'Object-dependent thoughts and psychological redundancy.' *Analysis* 51, 1–9.
- Russell B (1910–1911). 'Knowledge by acquaintance and knowledge by description.' *Proceedings of the aristotelian society, new series* 11, 108–128. [Reprinted in *Mysticism and logic*. London: George Allen and Unwin, 1917.]
- Russell B (1912). *The problems of philosophy*. Oxford: Oxford University Press.
- Russell B (1918). 'The philosophy of logical atomism.' *Monist* 28, 495–527. Reprinted in Russell B (1956). *Logic and knowledge* London: George Allen and Unwin.
- Schiffer S (1978). 'The basis of reference.' *Erkenntnis* 13, 171–206.
- Searle J (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Searle J (1991). 'Response: reference and intentionality.' In Lepore & Van Gulick (eds.).
- Segal G (1989). 'The return of the individual.' *Mind* 98, 39–57.

Objectivity in Moral Discourse

M Timmons, University of Arizona, Tucson, AZ, USA

© 2006 Elsevier Ltd. All rights reserved.

Does moral discourse purport to be objective? If so, can its objectivist pretensions be justified? Roughly speaking, to say that some form of discourse is objective is to say that there is a single set of truths about whatever subject matter the discourse is about. We can make this rough characterization more precise by formulating a number of related theses – objectivity theses – that collectively capture this idea of objectivity:

Cognitivism: Declarative sentences of the discourse in question are used by speakers to make genuine assertions and function mainly to express the speaker's beliefs. Given that (sincere) belief and assertion aim at representing what is true, such sentences are capable of being true or false.

Truth: Some affirmative sentences of the discourse are true.

Independence: What makes some sentence of the discourse true is that it corresponds to some fact – where the existence and nature of the specific fact to which the sentence corresponds is independent of the attitudes and beliefs that individuals and groups have toward the fact in question. It is common to use the word 'stance' to refer to the beliefs and attitudes of individuals and groups, so using this terminology we may express the independence thesis by saying that for the discourse in question there is a stance-independent reality (realm of facts) that the discourse is about, and the facts comprising that reality are what make certain sentences of that discourse true.

Convergence: Ideally, use of the proper methods of inquiry (which may differ from discourse to discourse) may be expected to lead individuals (at least under suitably ideal conditions) to converge in a great many of their judgments about the subject matter of the discourse in question.

Discourse satisfying these four theses involves sentences that are (or may be) objectively true and hence we may say that the discourse itself is objective. Of course, it is possible for some realm of discourse to feature claims that **purport** to be objectively true but fail to be so because one or more of the above theses fail to hold in relation to the discourse in question – a possibility to which we shall return below.

Moral objectivism, then, at least as it is commonly understood by philosophers, is the view that all four theses hold in relation to moral discourse: not only does moral discourse purport to be objective, it satisfies the objectivity requirements in question and thus some moral judgments are objectively true. Sometimes this is put by saying that there is a single true morality. As we shall see, some philosophers have denied one or more of the four theses in relation to moral discourse and thus denied that morality is objective.

Determining whether moral objectivism is correct requires an examination of fundamental philosophical questions about the meaning, truth, and justification of moral judgments – questions central to that branch of ethics called 'meta-ethics.' But it also requires an examination of the very notion of objectivity. And here we find two models of objectivity. One of them is inspired by discourse about the physical world

(including ordinary nonscientific discourse about the world as well as scientific discourse) and which I will refer to as the model of ‘ontological objectivity’ – so called because the central idea is that there is an ontological realm of ‘really’ existing objects and properties that sentences from a particular discourse purport to be about and which serve to make true certain of those sentences. An alternate, more modest form of objectivity, which I will call ‘methodological objectivity,’ is less focused on matters of ontology and more focused on methods of reasoning that govern the discourse in question.

In what follows, we begin with a brief characterization of the ontological model of objectivity and then proceed in the next four sections to consider some of the evidence for and controversy about whether moral discourse is ontologically objective. After reviewing the pros and cons of the claim that moral discourse is ontologically objective, we turn to the second model of objectivity – a model that promises to make sense of objectivity without ontology.

Ontological Objectivity

In meta-ethical discussion over the objectivity of moral discourse, discourse about ordinary common-sense objects and their properties as well as scientific discourse – often referred to as descriptive discourse – is taken to be paradigmatic of objective discourse. Roughly, the idea is that descriptive discourse is (or at least purports to be) about a realm of objects and properties in certain combinations (facts) whose nature and existence is independent of our beliefs, conventions, attitudes, and other mental stances we might take toward those facts. Here, the emphasis is on considerations of ontology – on what exists – hence the label ‘ontological objectivity.’ Moreover, the ontological objectivist typically embraces a strong form of the independence thesis, viz.,

What makes some sentence true is that it corresponds to some sort of stance-independent fact – where the existence and nature of the fact in question is independent of the actual and ideal stances of individuals or groups.

The significance of this thesis of strong independence (SI) will emerge later on when we turn to methodological objectivity.

So the model of ontological objectivity (using descriptive discourse about objects and properties in the world as an example) clearly involves the first three objectivity theses: descriptive discourse serves to express beliefs (cognitivism), some of which are true (truth) when they correspond to

stance-independent facts (independence). Barring epistemological skepticism, we may also expect suitably motivated inquirers, using proper methods of inquiry, to converge in their beliefs about this realm of objective fact since their inquiries are being constrained by an objective, stance-independent reality.

Ontological Objectivity and Moral Discourse

Does moral discourse have the trappings of ontologically objective discourse? Indeed, is it a form of **descriptive** discourse – discourse about a special subject matter, but nevertheless a discourse that is properly interpreted as representing a realm of objective moral facts? Those who think so often point to a number of ‘markers’ – features that are deeply embedded in moral thought and discourse – that either reflect or at least seem to support the various objectivity theses. Let us consider some of them in more detail.

Grammatical and Semantic Markers

Perhaps the most obvious markers of objectivity concern matters of grammar and semantics.

O1: Moral sentences such as ‘John’s lying to Brenda was wrong,’ in which a moral term appears, are in the indicative mood and are used to make genuine assertions that express a (sincere) speaker’s beliefs.

O2: Moreover, because moral sentences are typically used to make assertions they are truth apt: they are candidates for semantic evaluation in terms of their truth and falsity. Sentences which predicate the truth or falsity of a moral sentence seem to make perfect sense (e.g., ‘It is true that John’s lying to Brenda is morally wrong,’ ‘It is false that contraception is morally wrong,’ and so on).

O3: Moreover, logically simple moral sentences (e.g., ‘It is wrong to steal’) can embed in all sorts of logically complex grammatical constructions, most notably truth-functional and quantificational constructions (e.g., ‘If it is wrong to steal, then it is wrong to encourage others to steal.’) It would seem that such logically complex sentences have truth values that are determined by the truth values of their simpler constituents.

Ontological Markers

The following two markers have to do with the ontological status of the subject matter of moral discourse.

O4: Moral terms that appear in moral sentences (e.g., 'good,' 'right,' 'virtuous,' and their opposites) seem to be used to denote properties that are (or may be) possessed by items of moral evaluation. For example, to say of an action that it is wrong appears to attribute the property of wrongness to the action. Furthermore, some moral sentences are used with the apparent intention of picking out moral properties and talking about them, as when one utters the sentence 'It was the badness of the practice of American slavery that eventually led to its abolishment.'

O5: Moral truth and moral error are genuine possibilities. When we disagree over some moral issue, we take ourselves to be engaged in a genuine dispute where not all parties to the dispute are correct in their moral convictions. What is apparently being assumed in such disagreements is that there is a fact of the matter about the issue in dispute – that there is an objective truth about the matter. This makes it natural to think that the truth (or falsity) of a moral sentence is a matter of its correspondence (or lack of correspondence) to some moral fact. The putative fact that an action is wrong (supposing it is) would seem to be what makes true a sentence expressing this fact. Moral error results when what one says or believes fails to properly record moral reality.

Epistemological Marker

Finally, the following marker has to do with matters of justification and knowledge.

O6: We dispute moral claims by offering reasons for their truth or falsity and thus take such claims as being susceptible to justification. One can be justified or unjustified in claiming that John's action was morally wrong depending on one's evidence for the claim – evidence that can be offered, disputed, and discussed with others. Moreover, in some cases we at least suppose that we have some moral knowledge. Surely we take ourselves to know that torturing puppies for fun is morally wrong. Since justification and knowledge in ethics (as elsewhere) seems to require that there be reliable methods of inquiry suitable to the form of inquiry, one would expect there to be a moral methodology that would enable inquirers using it to reach a high level of interpersonal convergence in moral belief.

Thus, moral discourse, at least at first glance, seems to possess all of the characteristic marks of objectivity – indeed ontological objectivity. Many philosophers go one step further and claim that moral discourse is genuinely ontologically objective

and that moral discourse is a kind of descriptive discourse whose primary function is to represent a realm of moral facts.

But ontological moral objectivism is controversial: we find a variety of meta-ethical positions that deny one or more of the four objectivist theses and among moral objectivists we find a variety of specific meta-ethical positions that differ over questions of semantics, metaphysics, and epistemology in relation to moral discourse. Let us now briefly survey some of these positions.

Moral Realism

One very straightforward way to accommodate O1–O6 is simply to affirm the four objectivist theses and, in particular, the idea that there is a realm of moral properties (and moral facts) that are strongly independent of our attitudes and beliefs about them and which serve as an objective basis for the truth and falsity of moral sentences. This meta-ethical position is known as moral realism and the task of this kind of realist view is to explain what sorts of fact serve as truth makers for moral judgments and how it is possible to have justified belief and perhaps knowledge of moral truths. Ontological naturalists, who are moral realists (Sturgeon, 1984; Railton, 1986, 1996; Boyd, 1988; Brink, 1989; Bloomfield, 2001), claim that moral properties and facts are identical to certain natural properties and facts – properties and facts that are the proper subject matter for empirical science to investigate. In the history of meta-ethical inquiry, we find various attempts by philosophers to 'reduce' moral properties and facts to some species of biological, psychological, or sociological fact. For instance, Spencer (1895) held that the property of being better than, which he took to be basic in ethics, could be reduced to the property of being more highly evolved. Ontological nonnaturalists in ethics (Moore, 1903; Ross, 1930; Shafer-Landau, 2003) hold that moral properties and facts are of a special nonnatural kind; that rightness, for example, is not identical to any natural property, but is nonetheless a real property.

Naturalist realists face the challenge of explaining why some specific set of natural properties and facts is identical to moral properties and facts. This in turn would seem to require a plausible semantic account of moral judgments that fits with naturalist realism, something which has proved difficult to do. Nonnaturalists, by contrast, propose to bloat our ontological commitments: in addition to being metaphysically committed to those entities in the natural

world that common sense and natural science commits us to, the nonnaturalist wants to add a realm of nonnatural properties and facts. Such metaphysical extravagance has been met with strong resistance from philosophers who think that such properties and facts are mysterious metaphysical accretions and that there are serious epistemological worries accompanying any such view.

Denying Ontological Objectivity in Moral Discourse

Some philosophers deny moral objectivism because they deny the thesis of cognitivism in relation to moral discourse. They claim that although the surface grammar strongly suggests that moral judgments express beliefs and thus are used to make truth-apt assertions, nevertheless all of this is misleading. Rather, according to the meta-ethical position known as noncognitivism (Stevenson, 1937, 1944; Ayer, 1946) and more recently as expressivism (Gibbard, 1990, 2003; Blackburn, 1993, 1998; Horgan and Timmons, 2006), moral judgments are not really in the business of expressing beliefs that purport to represent or describe moral facts; rather they instead function to express some noncognitive attitude toward the object of evaluation. Thus, for instance, according to emotivism (one kind of noncognitivism), uttering a sentence such as 'Abortion is wrong' really functions primarily to **express** one's negative feeling toward abortion and is roughly equivalent to saying, 'Abortion: boo!' So, the noncognitivist denies both the theses of cognitivism and truth (in relation to moral discourse) and hence denies the independence and convergence theses as well. What makes this view fairly radical is that it distinguishes between the surface trappings of moral discourse and its true, deep semantic working, and claims that a proper semantic interpretation of the discourse reveals that it really does not even **purport** to be objective, that, to repeat, moral sentences are not really used in thought and discourse to express beliefs capable of being true or false.

A slightly less radical view, the error theory (Mackie, 1977), maintains that although moral discourse is properly interpreted as at least purporting to be objective, its various objective pretensions result from massive error. More precisely, the error theorist accepts the thesis of cognitivism in relation to moral discourse, but denies the thesis of independence and so denies the claim that affirmative moral sentences are ever objectively true. This position is analogous to atheism. That is, an atheist holds that religious discourse involving affirmative claims about God or gods purports to be

objectively true, but, claims the atheist, there are no theological facts of the relevant sort that can make true affirmative claims about God or gods, and so no such claims are objectively true. The error theorist in ethics is saying something similar. The moral judgment 'repaying one's debts is morally right' purports to attribute the property of moral rightness to the activity of repaying debts, but since there are no such moral properties this sentence and all affirmative moral sentences (taken literally) are false.

Finally, versions of moral relativism (Harman, 1984; Wong, 1984) typically affirm the theses of cognitivism and truth, but maintain that moral truth is not independent of the attitudes of individuals and groups. Rather, for a typical relativist, what makes a moral claim true is that some group accepts some set of moral principles and these principles are the very standard (for that group) that determine (together with nonmoral facts) which further moral sentences are true. The idea is that the sentences expressing moral principles are counted as true in virtue of being accepted by some group and the more specific moral sentences that (together with nonmoral facts) follow from these principles express derivative moral truths. So, if some group accepts as a basic moral principle that eating meat is morally wrong, then the sentence 'Eating meat is wrong' is true – relative to their moral outlook. If some animal (or its remains) is as a matter of fact meat, then this fact together with the general moral principle in question implies specific moral truths about eating this or that animal or its remains. Now, if some other individual or group does not have any such principle against meat eating, or has a principle that requires eating meat, then the sentence 'Eating meat is wrong' is false – relative to their outlook. Thus, for the relativist, conflicting moral sentences may be equally true.

There is a variety of ways to develop this basic relativist idea, but all of them are committed to denying the idea that there is some realm of moral fact that is independent of the stances of individuals and groups; in short, for the moral relativist, there is no single true objective morality.

Noncognitivism, the error theory, and relativism have their able defenders, but of course embracing one of these positions means having to claim either that at least some of the various objectivist markers (O1–O6) are illusory – not really characteristic of the discourse – or that they are genuine but deeply error ridden. Such views strike many philosophers as meta-ethical options we should try to avoid in making sense of moral thought and discourse (Wright, 1992; Timmons, 1999).

Methodological Objectivism

In light of the metaphysical and epistemological problems that beset realist (ontological) accounts of moral objectivity, some philosophers are attracted to a different model of objectivity – a model that looks to the realm of mathematics and logic for a way of understanding how moral discourse can be objective even if there are no dedicated moral properties and facts (of the sort associated with ordinary descriptive discourse) that would serve as truth-makers for moral sentences. Such meta-ethical views are often called ‘constructivist.’ A central idea behind versions of moral constructivism, then, is that just as we need not suppose that there is a mystical realm of numbers and mathematical relations that make certain mathematical sentences true, so we need not make any heavy-duty ontological assumptions in making sense of the apparent objectivity of moral discourse (Putnam, 2004). For both sorts of discourse – mathematical and moral – so long as there are **methods** of reasoning governing these types of discourse that would lead properly motivated individuals to converge in their views on a large enough number of claims within those areas, then we have a proper basis for affirming the objective pretensions of both realms. So, in contrast to the ontological model of objectivity that emphasizes an SI thesis, the present model takes the thesis of convergence as central to moral objectivity. Let us explore this model in a bit more detail.

Suppose, then, that associated with moral discourse is a method of moral thought and reasoning that would (if properly applied) ideally lead to interpersonal convergence over a wide range of moral issues; in other words, suppose the convergence thesis holds for moral discourse. Then, according to this line of thought about objectivity, we would have the materials to vindicate the other objectivity theses – cognitivism, truth, and a modest form of independence. How might this view be developed?

The general constructivist idea is that there are better and worse methods of moral thinking – methods that involve constraints on moral deliberators and on the circumstances in which they deliberate. For instance, according to one variety of constructivism, called the ‘ideal observer’ theory (Firth, 1952; Carson, 1984), the proper method of moral thinking requires that moral deliberators thinking about a certain moral issue must for example have certain intellectual virtues and sufficient knowledge of non-moral facts to be in a position to arrive at moral verdicts that have the status of being true. But what is crucial to the ideal observer view and constructivist

views generally is that it is the attitudes of ideal observers that constitute or make true certain moral judgments. So, although the constructivist accepts the independence thesis (in addition to cognitivism, truth, and convergence), she embraces the following modest form of this thesis in contrast to the strong form embraced by the ontological objectivist:

Although the truth of a moral sentence is independent of the stances of actual individuals and groups, moral truth is ultimately constituted by the stances of ideal individuals or groups – call them ‘ideal stances.’

Granted, some moral constructivists introduce ontological talk of moral properties and facts at this point (Rawls, 1980; Scanlon, 1998), understanding such properties and facts to be ‘constructions’ grounded in facts about moral convergence. But a moral constructivist need not make these ontological claims; she can rest content in emphasizing the importance of there being a method of moral inquiry that would lead all or most ideal agents (agents who are engaged in moral thinking under ideal circumstances) to converge in a great many of their moral convictions (Smith, 1994). What the constructivist hopes to do, then, is develop a meta-ethical view that satisfies the four main objectivist theses without (apparently) having to embrace heavy-duty ontological assumptions, and thereby vindicate the objectivity of moral discourse. In short: moral objectivity without (moral) ontology.

Of course, any constructivist view carries the burden of spelling out a moral methodology or process of moral reasoning that would lead ideal moral inquirers to converge in their moral views. The task of doing so poses a seeming dilemma for the moral constructivist (Timmons, 2004). On the one hand, if the favored moral methodology that is to be a basis for moral truth is characterized in morally neutral terms, then the constraints on such reasoning will be insufficient to lead those following the method to reach the level of convergence in their moral verdicts that a vindication of objectivism requires. This would mean, according to the constructivist, that there is no objective moral truth about those moral matters regarding which moral methodology fails to yield convergence under ideal conditions. Hence, the view would be committed to a kind of moral error theory – there is no fact of the matter in relation to a great many moral issues, contrary to what we suppose. On the other hand, if, in order to make the moral methodology yield substantive moral verdicts, one characterizes the methodology making use of moral assumptions, then relative to one set of

moral assumptions the method will yield one set of moral 'truths,' but relative to a competing set of moral assumptions the method will yield a different and perhaps conflicting set of moral 'truths.' But this commits the constructivist to moral relativism, which, as we have seen, is at odds with the idea that moral discourse is objective – that with respect to a large number of moral issues there is a single truth of the matter. Avoiding the error theory and relativism and thereby preserving objectivity is thus the main challenge for the moral constructivist.

Conclusion

The issue of moral objectivity, involving, as we have seen, semantic, metaphysical, and epistemological issues, is the central topic of meta-ethics. About the objectivity of moral discourse there are two questions to be distinguished. First, what sorts of objective features or markers are genuinely possessed by moral discourse? Second, assuming that moral discourse does genuinely possess such markers, how should they be interpreted? In particular, do they require, for their vindication, some version of moral objectivism? In response to these questions, we have surveyed a number of currently debated meta-ethical options, including two conceptions of objectivity.

Bibliography

- Ayer A J (1946). *Language, truth and logic* (2nd edn.). New York: Dover.
- Blackburn S (1993). *Essays in quasi-realism*. Oxford: Oxford University Press.
- Blackburn S (1998). *Ruling passions*. Oxford: Oxford University Press.
- Bloomfield P (2001). *Moral reality*. Oxford: Oxford University Press.
- Boyd R (1988). 'How to be a moral realist.' In Sayre-McCord G (ed.) *Essays on moral realism*. Ithaca, NY: Cornell University Press.
- Brink D O (1989). *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.
- Carson T (1984). *The status of morality*. Dordrecht: Reidel.
- Copp D & Zimmerman D (eds.) (1984). *Morality, reason, and truth*. Totowa, NJ: Rowman and Littlefield.
- Dworkin R (1996). 'Objectivity and truth: you'd better believe it.' *Philosophy and Public Affairs* 25, 87–139.
- Firth R (1952). 'Ethical absolutism and the ideal observer.' *Philosophy and Phenomenological Research* 12, 317–345.
- Gibbard A (1990). *Wise choices, apt feelings*. Cambridge, MA: Harvard University Press.
- Gibbard A (2003). *Thinking how to live*. Cambridge, MA: Harvard University Press.
- Hare R M (1993). 'Objective prescriptions.' *Philosophy* 35, 1–17.
- Harman G (1984). 'Is there a single true morality?' In Copp D & Zimmerman D (eds.), 27–48.
- Horgan T & Timmons M (2006). 'Cognitivist expressivism.' In Horgan T & Timmons M (eds.) *Meta-ethics after Moore*. Oxford: Oxford University Press.
- Leiter B (ed.) (2001). *Objectivity in law and morals*. Cambridge: Cambridge University Press.
- Mackie J L (1977). *Ethics: inventing right and wrong*. New York: Penguin.
- McNaughton D (1988). *Moral vision*. Oxford: Blackwell.
- Moore G E (1903). *Principia ethica*. Oxford: Oxford University Press.
- Putnam H (2004). *Ethics without ontology*. Cambridge, MA: Harvard University Press.
- Railton P (1986). 'Moral realism.' *Philosophical Review* 95, 163–207.
- Railton P (1996). 'Moral realism: prospects and problems.' In Sinnott-Armstrong W & Timmons M (eds.) *Moral knowledge? New readings in moral epistemology*. Oxford: Oxford University Press. 48–81.
- Rawls J (1980). 'Kantian constructivism in ethics.' *Journal of Philosophy* 77, 512–572.
- Ross W D (1930). *The right and the good*. Oxford: Oxford University Press.
- Scanlon T (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Shafer-Landau R (2003). *Moral realism: a defence*. Oxford: Oxford University Press.
- Smith M (1994). *The moral problem*. Oxford: Blackwell.
- Spencer H (1895). *The principles of ethics* (2 vols). New York: Appleton.
- Stevenson C L (1937). 'The emotive meaning of ethical terms.' *Mind* 46, 14–31.
- Stevenson C L (1944). *Ethics and language*. New Haven: Yale University Press.
- Sturgeon N (1984). 'Moral explanations.' In Copp D & Zimmerman D (eds.), 49–78.
- Timmons M (1999). *Morality without foundations: a defense of ethical contextualism*. New York: Oxford University Press.
- Timmons M (2004). 'The limits of moral constructivism.' In Stratton-Lake P (ed.) *On what we owe to each other*. Oxford: Blackwell. 90–122.
- Wong D (1984). *Moral relativity*. Berkeley/Los Angeles: University of California Press.
- Wright C (1992). *Truth and objectivity*. Cambridge, MA: Harvard University Press.

Objects, Properties, and Functions

R J Stainton, University of Western Ontario, London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

One of the most fundamental metaphysical distinctions is that between objects and properties. Given its depth and importance, it's perhaps unsurprising that though the distinction can be given in a canonical form, it also admits of some puzzling complexities.

By way of a first pass, let's consider some examples of objects and of properties and try to extract some hallmarks – linguistic and otherwise – of each. Examples of objects include my cat, you, the tree outside my window, and the planet Earth. These four are all physical objects. There are mental objects as well – which, depending upon whether the mental reduces to the physical, may or may not be a subvariety of physical objects. Mental objects include my wish to write a book, the pain I felt just now, my dream last night. So much for examples. Some hallmarks of objects include that they are individual things, which things (at least in the simple cases) have a specific location in space and time. What's more, objects can be counted. And, in counting, one can always ask: Is this object the same individual as that one, so that it should be counted just once, or are there two individuals here? (E.g., if Samuel Clemens is the same person as Mark Twain, then when Clemens is alone in the room, there aren't two individual people there: Clemens *and* Twain. But if Bill Clinton and Tony Blair are the sole people in the room, two people are therein.)

Examples of properties – also called 'attributes,' or "qualities" – include being red, being from Canada, being painful, knowing Charles Darwin personally, weighing exactly 16 trillion kilograms, and having belonged to Abe Lincoln. Hallmarks of properties include *not* being individual but rather being general, and being located not in one place but at various locations at once or at none. Thus the property of being from Canada is instantiated in many objects, in many places; and it might be that the property of knowing Charles Darwin personally is instantiated by no object at all these days. As for counting, while one can count *instantiations* of being red – i.e., count how many red objects there are – it doesn't really make sense to count being red itself. Another hallmark of properties, as opposed to objects, is that a given property typically has contrary or opposite properties. For instance, being green is contrary to being red, being from Spain is contrary to being from Canada, being pleasurable is contrary

to being painful. (Contrast objects: What is the contrary of Abe Lincoln? What object do you get when you negate *him*? Or, again, what do you get when you negate my window? These questions don't even seem to make sense.)

There are also *linguistic* hallmarks of the object-versus-property divide. The linguistic correlates of objects are proper names ('Rob'), noun phrases ('the planet Earth'), pronouns ('me,' 'he,' 'she,' 'it'), and demonstratives (like 'this' and 'that'). The linguistic correlates of properties, by contrast, are adjectives ('red'), verb phrases ('weighs 16 trillion kilograms,' 'is from Canada') and, as above, gerundives ('belonging to me'). (In formal logic, the divide corresponds to that between constants and variables [often symbolized by lowercase letters from the beginning of the alphabet for constants, and lowercase letters from the end of the alphabet for variables: constant *a* or variable *x*] on the one hand and predicates [often symbolized by capital letters from the middle of the alphabet: F, G] on the other.) Another linguistic hallmark is this. If, following Peter Strawson (1971), we think about the different actions typically performed in speaking a sentence – for instance in saying "That cat weighs 18 pounds" – we may note that "That cat" is used for one kind of job, while "weighs 18 pounds" is used for a quite different job. "That cat" is used to *refer to*, or *talk about*, the cat. In contrast, "weighs 18 pounds" is used to *attribute* something of the cat, to *say of* it that it weighs 18 pounds. A final linguistic hallmark, then, is that in speaking, we refer to objects, but we attribute properties.

So much for the first pass at objects versus properties. The divide gets muddy, however, when we leave the canonical examples. That's because there are things that are objectlike in some respects but propertylike in others. These hard-case objects come in at least two varieties: *abstracta*, like numbers, and *kinds*, such as gold and horse. One obvious problem with such objects concerns their location. Being abstract, the square root of seven doesn't seem to have any location. The same holds for kinds: gold has multiple locations. Moreover, there are contraries to gold: silver, uranium, etc. Yet abstracta and kinds are objectlike as well. Gold *has* certain properties: it boils at 2966 °C and melts at 1063 °C. So, it isn't itself a property. And gold, along with the square root of seven, is denoted by a name; and, in speech, such things are referred to, not attributed. So, kinds and abstracta seem to exhibit some of the hallmarks of objects and some of the hallmarks of properties. At bottom, the difficulty is that such objects have instances, rather than merely being instances.

To take another case, consider the novel *War and Peace*. Some will say that it's another abstract object: it itself has no location, though its instances (i.e., copies of the book) do. But one could also take *War and Peace* to be like gold, which is located in many places at once. Whichever way one goes, it too has instances, rather than merely being an instance; and so it does not have one unique location. See **Figure 1** for a summary.

One last hard case. It seems that we can form names and descriptions associated with properties. Thus, consider 'the property of being red.' This description has all the linguistic hallmarks of an object, in terms of both its form and the job it is used for. More than that, it has some of the nonlinguistic hallmarks too. We can ask whether what 'the property of being red' stands for is identical to what 'the property of being green' stands for. (Obviously the answer is no.) We can count such things. And so on. What one typically says is that 'the property of being red,' 'that property,' 'my favorite color,' and so on stand for objects that are associated with properties but do not themselves stand for properties. (Of course, this leads to the odd result that 'The property of being red is not a property' is not actually false, since according to the view under consideration, the first five words of this sentence *stand for an object*, so the sentence says of this object that it's a property – which it isn't! See Frege 1891, 1892 for a detailed discussion.)

One could continue pursuing such complexities at enormous length. For present purposes, it's enough to notice that there seem to be various kinds of objects. And several of them are propertylike in certain respects. It is in part because of the diversity of objects, and their affinities to properties, that it's hard ultimately to distinguish objects from properties.

In light of these complexities, one might wonder: Why are properties required at all? Why not just have yet another variety of object – namely, those objects that don't merely have instances (as with kinds and

abstracta) but that can actually apply to other objects? Indeed, why not have the world be exhausted by its objects? Thus, on this suggestion, being from Canada would simply *be* a fourth kind of object beyond instances, kinds, and abstracta. In response, I want to give one reason that one seemingly cannot do away with the object/property distinction altogether, reducing everything to objects. It's a reason that goes back to Antiquity, and it has to do with combining in predication.

Patently, some objects can't combine predicationally at all: Rob and the tree can't be fused together into a complex fact. At best they can be combined by smashing them together or putting one inside the other or some such. But, so the 'properties are objects' idea goes, there are some objects that can combine in predication; there are certain objects that apply to others. Take Rob Stainton and being Canadian, understood, as per the proposal, as two different kinds of objects. These 'objects' combine together, in predication, in the sense that the former exhibits the latter: I am Canadian. The difficulty, however, is this: Not all objects actually so combine, even when one of them is of the supposed 'applying kind.' For instance, Rob Stainton doesn't in fact combine with being Algerian, and George Bush doesn't combine with being Canadian. So, what holds the two types of supposed objects together when the 'applying object' really does combine with the instance one? A natural thought is that the relation 'exhibiting' puts together the objects that really go together, and not the others: 'exhibiting' fuses Rob with being from Canada, but it does not fuse Rob with being from Algeria. However, if *all* things are objects, then relations – including 'exhibiting' – are objects too. And if, as we've been seeing, objects need something to put them together to distinguish those where the combination does really happen from those where it doesn't, 'exhibiting' needs something to put *it* together with – e.g., Rob and being Canadian. What, then, puts together these three objects in predication such that it doesn't fuse 'exhibiting,' Rob and being Algerian? We are faced with a choice: either there is something that is not an object and that puts objects together in predication, so that not everything is an object, or there is an infinite regress. Since the latter is absurd, we conclude that some things aren't objects. In particular, properties are non-objects, in the sense that they go together with objects in a way that two objects cannot: nothing is required to 'fuse' an object with a property. That, it is often said, is why we need some kind of object/property divide, even if it is hard to lay out.

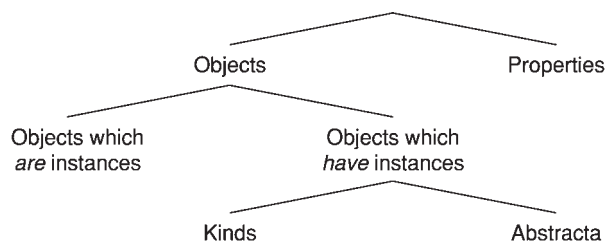


Figure 1 Varieties of objects.

An interesting upshot, of course, is that not all that is, is an object. Even more interesting, if anything we can refer to is an object, then if properties are not objects, we simply cannot refer to them. We can't talk about them, no matter how hard we try. At best we can refer to their associated surrogates, as in the previous example of 'the property of being red' – which surrogates are objects, not actually properties.

Let's turn now to the relationship between properties and functions. What I hope to explain is how properties are, in a certain sense, equivalent to a peculiar kind of function: namely, a function whose output is a truth value. (See Frege 1879, 1891, 1892.) Now, mathematical functions are familiar: they take numerical inputs and yield numerical outputs. For instance, (1) is a function that takes zero as input and yields four as output, and takes one and yields nine as output, and so on.

$$(1) y = (x + 2)^2$$

But the notion of function can be generalized so that the inputs are not just numbers but objects of all kinds. To understand the broadening of the inputs, consider (2), which is a function that takes an individual as input and yields his or her father as output.

$$(2) y = \text{the father of } x$$

If you take me as input, the output is Keith Stainton. One could also have a function whose input was a country and whose output was the language most widely spoken in that country. If Canada were taken as input, this function would output English. It should be clear that for any kind of object, we can cook up a function into which it can be input: trees, windows, universities, etc.

So, inputs to functions can be generalized. Moreover, their outputs needn't be numbers, either. Consider, in particular, a function whose output is either TRUE or FALSE (or yes/no, or 1/0, or any other pair of binary values). Specifically, consider a function that outputs TRUE if the object input meets the condition specified and that outputs FALSE if it doesn't. For instance, consider the function such that its output is TRUE if x is from Canada and its output is FALSE otherwise. This is the key point: such functions are interchangeable with the properties. To the property of being from Canada, there corresponds f , the function from objects to truth values, such that f outputs TRUE when the object is from Canada and outputs FALSE otherwise. (Gottlob Frege called

such peculiar functions 'concepts'; Bertrand Russell called them propositional functions.)

We saw above that objects and properties have different linguistic correlates. Given that propositional functions are interchangeable with properties, it is unsurprising that a similar point applies to them. Specifically, the correlate of a verb phrase or an adjective will be a function from an object to a truth value. (In terms of Montague grammar, the correlate of objects will be expressions of semantic type $\langle e \rangle$, while the correlate of (first-order) functions will be expressions of semantic type $\langle e, t \rangle$.) And, as before, puzzles arise about whether functions are really so different from objects. At first the distinction is clear: objects are the inputs and outputs, while functions operate on them. But it seems that we can objectify functions and talk about them. This I did above: I talked about the function from things that are from Canada to TRUE. Put another way, there are functions whose inputs are functions, and functions whose outputs are functions. What's more, there are alleged objects, such as the kind gold, that seem to be true of instances; hence, they could be taken to be a propositional function, rather than as objects. But then are they just another kind of object? The answer here, again, is that there can't only be objects: we'd need something to combine 'ordinary objects' with 'function objects' that wasn't itself of either kind, or we'd have a regress. What goes for properties goes for propositional functions: there may be various kinds of objects, some of which exhibit functionlike characteristics, but we cannot do away with genuine functions altogether.

See also: Metaphysics, Substitution *Salva Veritate* and the Slingshot Argument.

Bibliography

- Frege G (1879). *Begriffsschrift*. Halle: L. Nebert.
- Frege G (1891). 'Function and concept.' In Beaney M (ed.) *The Frege reader*. Oxford: Blackwell. 130–148.
- Frege G (1892). 'On concept and object.' In Beaney (ed.) 181–193.
- Russell B (1905). 'On denoting.' *Mind* 14, 479–494. In Marsh R C (ed.) *Logic and knowledge*, London: Unwin Hyman.
- Strawson P F (1971). *Logico-linguistic papers*. London: Methuen.

Ordinary Language Philosophy

P Snowdon, University College London, London, UK

© 2006 Elsevier Ltd. All rights reserved.

‘Ordinary language philosophy’ is the name given to some conceptions of philosophy, which emanated, after the Second World War, from J. L. Austin in Oxford and from Wittgenstein in Cambridge, and which championed the importance for philosophy of attention to ordinary language. The name itself, the origin of which is not recorded, has perhaps been more used by opponents of these ideas than their friends and so has, to some extent, become a term of mild abuse.

Language and Philosophy

Language itself has always been of interest to philosophers. Our ability to understand and use language is such an obvious feature of our cognitive capacities that any attempt to describe and explain cognition must consider language. In the early modern period such philosophers as Locke and Berkeley provided theories of the nature of language, trying to account principally for its semantic properties. Their thinking, however, was dominated by a belief in the priority of thought, conducted according to them in a prelinguistic medium of ideas, over language. Language was regarded as a medium for public expression of private thoughts, and its interpretation was viewed as making words public ‘signs for’ private ideas. For them the problem, which generated considerable disagreement, was whether the seemingly manifest differences between different parts of language, between, say, names and adjectives, required the postulation of different sorts of ideas. This was, of course, a concern about the nature of ordinary language. By the beginning of the 20th century, a more comprehensive philosophical engagement with the nature of language emerged in the writings of Frege, Wittgenstein, and Russell. This engagement was inspired by the emergence of formal logic, with its semantic categories, and it was not driven by any prior commitment to the priority of thought over language. There was a real concern with a wide range of different constructions in ordinary language. Thus began what became a, perhaps the, central branch of philosophy, the philosophy of language. Of course, this concern with language was a concern with ordinary language, amongst other sorts of language.

Ordinary Language Philosophy

The name ‘ordinary language philosophy’ does not stand for this general theoretical concern in

philosophy with the nature of ordinary language or for any particular theory about ordinary language. Rather, it stands for two linked approaches to, or theories about *philosophy* itself. The approaches shared the conviction that the key to, or at least a necessary condition for, finally solving philosophical problems lies in some sort of detailed attention to ordinary language. This common theme led philosophers at the time to think that there was a single movement, representing a revolutionary approach to philosophy, which held out hope of laying recalcitrant philosophical problems to rest. However, as we shall see, it is an illusion to think there is a common movement here.

The Oxford Version: Austin

The Oxford ordinary language movement was led by, or inspired by, J. L. Austin, who became White’s Professor of Moral Philosophy at Oxford in 1956. Austin himself did not publish much during his lifetime, but established his leadership within a group of highly talented philosophers in Oxford, including Grice, Strawson, Urmson, and Warnock, partly due to his seniority, but also to his extraordinary critical intelligence and personality. Austin had a deep interest in language and wrote about what he called performative utterances, that is utterances such as saying, in the right circumstances, ‘I name this ship *Victory*,’ whereby the act of speech itself accomplishes what it ascribes, in this case conferring a name on a ship. He developed a more general theory of the acts we perform by speech in the William James Lectures he gave at Harvard in 1955, published under the title *How to do things with words*. In these he criticizes his earlier views on performatives and then introduces the famous distinction between locutionary, illocutionary, and perlocutionary acts, which are categories for different types of acts we do in or by using language. Austin also presented his own theory of truth. These publications contain Austin’s theories of language and of some of its properties.

Austin, however, thought that, in general, the right method when doing philosophy is to “proceed from ordinary language, that is, by examining *what we should say when*, and so why and what we should mean by it” (Austin, 1961: 129). In the most systematic presentation of his views, in his article ‘A plea for excuses,’ Austin recommended this for three reasons. The first is that philosophers need to ensure their own language is clear; the second is that we need to prize words apart from the world so that we can look at the world without blinkers; finally, ordinary language

contains many distinctions and subtleties that have withstood the tests of time and so must represent a subtle and plausible way of thinking about things, particularly things in normal life. Austin added that his view is that ordinary language will be the 'first word' when doing philosophy, though it need not be the 'last word.' He recommended that one should regularly consult a dictionary to get the sense of the categories and words related to a certain area or domain. Austin himself spent considerable effort plotting the distinctions that ordinary language categories embody. For example, what is the difference between a tool and a utensil? What is the difference between doing something willingly and voluntarily? Questions such as these were studied in Austin's famous Saturday morning discussion group in Oxford. It is indeed a reasonable conjecture that one thing that appealed to Austin about pursuing such questions is that they can be answered by cooperation within a group, amongst whom agreement can be reached, in sharp contrast to what has traditionally been the very solitary methods of philosophy with its continual disagreements.

Some Questions about Austin's Approach

There is, surely, no doubt that plotting the incredibly subtle distinctions of ordinary language is of interest in itself, and it certainly fascinated Austin. The main question that arises, of course, is what is its relation to, or value for, philosophy. The questions of philosophy are characteristically highly general, such questions as 'Is the will free?', and 'Do we know about the external world?' How, then, does a study of the subtle and rather specific categories of ordinary language help with them? The major weakness of Oxford-style ordinary language philosophy is that it is not based on any theory of the nature of philosophical questions that explains and justifies the relevance of a study of ordinary distinctions to it. In fact, it would have been somewhat anathema to the mind set of such ordinary language philosophers to have a general theory of philosophy, but it meant that its relevance remained obscure.

When Austin discussed philosophy head on, as he did in *Sense and sensibilia*, criticising the sense datum approach of Ayer, he noted that philosophers employ expressions in a way which does not fit their ordinary use. For example, he pointed out that the central terms 'direct perception' which often figure in the very formulation of the problem are used in a special way (see Austin, 1962: 14–19). It is a mistake though to take this discrepancy as a basis for criticism of the talk by the philosophers. It merely reveals that the philosopher's use is technical and in need of an

explanation. Austin also clearly thought that the categories of philosophers are far more general than those which are of normal employment. But, again, this fact, in itself, does not amount to a criticism of those general categories.

If this assessment of the gap at the core of Austin's approach is correct, it explains why the approach, despite the fascinations of its ordinary language investigations, simply withered in its appeal to those engaged by philosophical questions. A crucial moment was the publication in 1959 of Strawson's *Individuals*, with its self-avowed aim of doing highly general metaphysics.

The Cambridge Version: Wittgenstein

The second attitude to philosophy that was called 'ordinary language philosophy' is that of the later Wittgenstein, during his period after 1929 when he was primarily at Cambridge. The major presentation of the approach is in *Philosophical investigations*, published in 1953 after Wittgenstein's death in 1951. In the evolution of his later philosophy Wittgenstein was, in part, attempting to explain why his earlier, highly abstract and metaphysical philosophy, contained in the *Tractatus*, is wrong. In that earlier work, Wittgenstein attempted to describe the essential nature of contingent propositions, and the states of affairs that they report, basically by picturing the elements. There are also the tautologies of logic and mathematics. The pronouncements of philosophy itself seem not to belong to either category, and so the contents of the *Tractatus* were themselves inexpressible by its own standards. Philosophy itself was, therefore, a process you went through in order to realize you had to leave it behind. It is, as Wittgenstein put it, like a ladder you climb only to throw away, no longer needing it. In the process, though, something profound is shown, although not said. This is a highly unusual and paradoxical conception of philosophy.

In the later period Wittgenstein rejected his earlier ideas about the essence of saying, holding instead that speech and understanding are parts of human life, which may take many forms (as Wittgenstein might have said, 'We play many language games'), and which have no hidden common essence that a reflective process like philosophy can discern. However, he retained even in the later period an attitude to what philosophy is that is unusual and paradoxical. His theory is that there are what we might call two activities, which go under the name 'philosophy.' There is, first, traditional philosophy, which for Wittgenstein was often taken to be the *Tractatus*, according to which there are supposed to be distinctively

philosophical problems, concerned with the essence of things and our knowledge of the world, and theories are proposed as answers to these questions and rationally assessed. According to the later Wittgenstein, traditional philosophy is not actually engaging with real problems. They are, rather, pseudo problems. He suggested that engagement in traditional philosophy always stems from an initial mistake when one speaks, employing the words of ordinary speech, in a way which is not actually in accordance with how the term in question is really used. They start, that is, from linguistic mistakes. Wittgenstein's view is that these mistakes are not random, but stem from properties of the language we employ. For example, the fact that I can say both I have a pain in my foot and I have a bone in my foot may lead us to imagine that pains are special objects located in space in the same way as bones. He said, "A picture held us captive. And we could not get outside it, for it lay in our language and language seemed to repeat it to us inexorably" (Wittgenstein, 1963: sec. 115). Wittgenstein likened being misled to being tricked. "The decisive moment of the conjuring trick has been made, and it was the very one that we thought quite innocent" (Wittgenstein, 1963: sec. 308). This is a schematic characterization of what is, in many ways, itself a schematic conception of traditional philosophy.

What we might call real philosophy should, according to Wittgenstein, consist in doing what is necessary to remove the impulse to engage in pseudo theorizing. As he famously said; "What is your aim on philosophy? – To show the fly the way out of the fly bottle" (Wittgenstein, 1963: sec. 309). How is this to be done? Wittgenstein had a negative and a positive claim to make. The negative theme is that the impulse is not removed by offering a positive and novel theory about anything that philosophers have attempted to theorize about. That would be to engage in pseudo theorizing itself. The task, rather, is to remove the impulse, and it should be viewed as akin to exterminating a disease. "The philosopher's treatment of a question is like the treatment of an illness" (Wittgenstein, 1963: sec. 255). So the real philosopher should produce something like a pill which removes a headache. It is this idea that is meant when Wittgensteinians talk of philosophy as therapy. The positive recommendation is that the pill should be a reminder how language is actually used. "What we do is to bring back words from their metaphysical to their everyday use" (Wittgenstein, 1963: sec. 116). He added that "when I talk of language ... I must speak the language of every day" (Wittgenstein, 1963: sec. 120). Hence, the idea is that proper philosophy consists in reminding those in the grip of traditional

philosophy how we ordinarily speak. This, then, is the idea of Wittgensteinian ordinary language philosophy.

Some Questions About Wittgenstein's Approach

Of the many questions that Wittgenstein's later conception of philosophy prompts, I shall restrict myself to two. Does Wittgenstein's own practice actually fit it? It is fair to say that Wittgenstein is, in his later period, primarily a negative thinker, his primary aim being to show what is wrong with standard philosophical ideas about, for example, understanding, following rules, sensations, action, and necessity. However, he did not think that simply reminding people of what they say accomplishes such criticisms. Rather, his practice was to think himself into the views in a highly creative way to the point where it becomes clear that they represent illusions, but where it is also revealed why they might seem attractive. It is therefore highly misleading to summarize Wittgenstein's own practice as reducing to reminders of what ordinary people say. Second, although Wittgenstein officially seems to think that proper philosophy does not propose theories or add to understanding, he clearly advances theories or semi-theories himself. Thus, he positively links meaning and use (see Wittgenstein 1963: sec. 43), and he is often taken to be arguing for the clearly philosophical proposition that a private language is impossible (see for example, Wittgenstein, 1963: secs. 258–279). It is, therefore, colossally misleading to think that describing Wittgenstein as an ordinary language philosopher captures his real approach to philosophy.

The second question is whether there is any reason to accept Wittgenstein's official account of the nature of philosophy. There is no *a priori* demonstration by Wittgenstein that the massive variety of questions that traditional philosophers have grappled with are all pseudo problems, grounded at some point in a misuse of ordinary language. Indeed, that seems a remarkably implausible thought. To take an example, traditional philosophers often debated what it is to understand a word. Wittgenstein brilliantly criticized the most popular answer, but it does not follow that the question is misconceived and does not merit a theoretical answer. Moreover, Wittgenstein was, as we have seen, moved to indicate his own answer in terms of use and practice. It seems to me, therefore, that we should regard the attitude to philosophy in the later Wittgenstein as simply an ungrounded and independent commitment within his thought, which does not even fit his own practice.

Conclusion

Austin's own theories about language and knowledge and his critical discussions of the philosophy of perception have remained influential, but both the rhetoric about the centrality to philosophy of ordinary language and the practice of attending to it for the reasons that Austin gives have vanished. Wittgenstein's approach to philosophy is still accepted by some people, but for most the real interest of the later Wittgenstein lies in his discussions of such central concepts as understanding, meaning, rule following, privacy of experience, and so on, and not in his extreme attitude to philosophy.

However, the emphasis on attending to ordinary language lead to significant advances in the theory of language. When philosophers recommended studying ordinary language, they usually meant, or at least said, that it was to be done by asking whether we would or would not say a certain thing in a certain circumstance. So determining the verdict of ordinary language meant determining what would be said. The philosophical relevance of this was supposed to consist in its showing that a philosophical theory which affirmed that P is actually true in circumstance C, would be wrong if we would not say that P in those circumstances. This prompted Grice, who had been part of the circle around Austin, to reflect on the relation between what is true and what we would say. He noted that truth alone is not usually sufficient to lead to speech, but rather, speech is governed by principles, such as 'Be helpful' or 'Be informative.' He explained in terms of this theory how the saying of something can carry an implication that in no way corresponds to what the remark literally entails. For example, if I remark 'The Provost is sober today' my doing so carries the implication that he is normally not sober. Whether Grice's theory is correct is still under discussion, but the contrast between speech

and truth, which is damaging to the rhetoric of 'ordinary language philosophy,' has now become conventional wisdom in philosophy. This insight probably emerged when it did because of those tendencies known as 'ordinary language philosophy.'

See also: Cognitive Science and Philosophy of Language; Speech Acts.

Bibliography

- Austin J L (1961). *Philosophical papers*. Oxford: Clarendon Press.
- Austin J L (1962a). *Sense and sensibilia*. Oxford: Clarendon Press.
- Austin J L (1962b). *How to do things with words*. Oxford: Clarendon Press.
- Berlin I (ed.) (1973). *Essays on J. L. Austin* (essays 1 to 3). Oxford: Clarendon Press.
- Chappell V C (ed.) (1964). *Ordinary language*. Englewood Cliffs, NJ: Prentice-Hall. (An important collection, including papers by Austin, Mates and Cavell.)
- Fann K T (ed.) (1967). *Ludwig Wittgenstein*. New York: Dell Publishing.
- Fann K T (ed.) (1969). *Symposium on J. L. Austin* (Part 1). London: Routledge.
- Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press. (This includes Grice's James Lectures, and also two chapters, 9 and 10, discussing the nature of philosophy.)
- Hacker P M S (1996). *Wittgenstein's place in twentieth century analytic philosophy*. Oxford: Blackwell. (See esp. chaps. 5, 6, and 8 – an account of ordinary language philosophy from a point of view which is highly sympathetic to Wittgenstein.)
- Wittgenstein L (1961). *Tractatus Logico-Philosophicus*. Pears D F & McGuinness B F (trans.). London: Routledge and Kegan Paul Ltd.
- Wittgenstein L (1963). *Philosophical investigations*. Oxford: Blackwell.

Origin of Language Debate

C Neis, University of Potsdam, Potsdam, Germany

© 2006 Elsevier Ltd. All rights reserved.

Preliminaries

The origin of language belongs to the most prominent topics of linguistic discussions in the 18th century. There is no doubt that the debate on the origin of language peaked during a time of much attention to the

question of human faculties in general and to the investigation of the natural endowment of man in order to distinguish him from the realm of beasts or to promote man's emancipation from ecclesiastical dogma.

Of course, it was not only in the 18th century that the question of language origins became a subject of philosophical concern. Given the philosophical and anthropological relevance of the topic, it has always attracted the curiosity of scholars. Glottogonic tales of ancient times have influenced the development of

philosophical thinking on the origin of language. One of the best-known myths is the story of the Egyptian King Psammetichus I (died 610 B.C.), who wanted to find out which language was the oldest and ordered an experiment that involved two children to be raised in isolation. In the course of history, several sovereigns are said to have followed the example set by the Psammetichus experiment: King Frederick II the Stauffer (1192/1193–1150), the Scottish King James IV (1473–1515), and the Indian Mogul Akbar the Great (1542/1556?–1605).

Important benchmarks in the discussion of the origin of language are the foundational texts of Western culture: the Bible and the philosophy of ancient Greece.

In the Bible, there are two myths of crucial importance to the further development of the question: Adam's naming of the beasts and the following dialogue between Eve and the serpent, and the legend of the Tower of Babel which ascribes the emergence of the diversity of languages to an act of divine punishment. God vindicates his realm against the sinful attempt of Man to erect a tower reaching the sky by transforming the only language that united man until that time into an inconceivable variety of idioms.

These biblical legends raised the question of whether at the very beginning there was only one language (monogenetical hypothesis) or whether diversity reigned since language came into being.

Discussions on the Origin of Language in Classical Antiquity

In his dialogue *Cratylus*, Plato (428/427–349/8) (see Figure 1) offered important insights into the relationship between words and things (cf. Schmitter, 1987). The disputants Cratylus and Hermogenes tried to find out whether words are created by nature or made up by convention. Cratylus advocated the



Figure 1 Plato. Drawn by Klaus Froese.

natural origin of words, which was mainly based on principles of sound symbolism and imitation. The conventional origin of language was defended by Hermogenes, whose argumentation was essentially focused on the undeniable diversity of languages which seemed a suitable proof of the non-natural origin of words. A final solution to the question of the natural versus conventional origin of words was offered by Socrates (470/469–399), who was introduced as a mediator by Plato. Socrates was eager to show the complementary character of the two respective approaches.

The conventionalist approach had a deep impact on Aristotle's (384/383–322) conception of the relation between words and things. For Aristotle, this relationship was not granted by nature, but depended entirely on convention. It was by human consent, by convention, that an arbitrary relationship between words and things was established ("*non natura, sed ad placitum*").

An important contribution to the development of the discussion on language origins especially in the Enlightenment is the Epicurean conception. Epicurus (342–271 B.C.) gave his account of the origin of language in a letter to Herodotus in which he attributed the emergence of language to a natural necessity, thus rejecting the idea of a conventional language origin. Epicurean philosophy was made accessible to the Roman world by Titus Lucretius Carus (98–55 B.C.) who explained the origin of language in his *De rerum natura* as a result of the natural endowment of man. It was his own nature that forced man to utter the first sounds, and his elementary needs urged him to express his first vocalizations (*At varios linguae sonitus natura subegit mittere, et utilitas expressit nomina rerum*; Lucretius V: 1028–1029). For Lucretius, language was neither a divine gift nor the invention of an ingenious language-inventor, but the result of the physical predisposition of man who tried to express his desires in a way quite similar to animal communication.

Discussions on the Origin of Language in the Enlightenment

The Epicurean philosophy had a strong impact on the Enlightenment debate on the origin of language (cf. Gensini, 1999) and inspired particularly the thought of leading figures of French sensationalism such as Étienne-Bonnot de Condillac (1714–1780) (see Figure 2), Jean-Jacques Rousseau (1712–1778) (see Figure 3) or Julien Offray de La Mettrie (1709–1751). It is well-known that both Condillac and Rousseau heavily influenced the discussion of the origin of language which reached its climax in



Figure 2 Condillac. Drawn by Klaus Froese.



Figure 4 Descartes. Drawn by Klaus Froese.



Figure 3 Rousseau. Drawn by Klaus Froese.



Figure 5 Locke. Drawn by Klaus Froese.

the origin of language competition announced by the Berlin Academy for 1771 (cf. Aarsleff, 1974).

The origin of language discussion had already been decisively influenced in the 17th century by René Descartes' (1596–1650) (see **Figure 4**) account of the nature of mankind in comparison to animals in his *Discours de la méthode* (1637). For Descartes, the faculty of language was the distinctive criterion of the human species in contrast to the animal world, permitting man to maintain a position of absolute supremacy. The faculty of language was conceived by Descartes as an indicator of the existence of human understanding and the soul, in sharp contrast to the instinctive mechanisms of animals.

Descartes' position had provoked the critique of Condillac, who conceived only a gradual difference between humans and animals, thus undermining the Cartesian idea of an unbridgeable gap between men and beasts. Condillac's most influential work for the origin of language debate was his *Essai sur l'origine des connaissances humaines* (1746), in which he presented a post-diluvial view of the origin of language integrated into an epistemological framework, tracing

the genesis of the operations of the mind. The epistemological framework underlying Condillac's *Essai* was largely influenced by John Locke's (1632–1704) (see **Figure 5**) *Essay concerning Human Understanding* (1690), but Condillac aimed at the abolition of the Lockean dualism between sensation and reflection for the benefit of the introduction of a leading rôle of sensation as the only source of knowledge (cf. Ricken, 1984).

For his explanation of language origins, Condillac imagined a couple of infants of both sexes, living isolated in a desert just after the great flood. Tracing their first steps toward the invention of language, Condillac described a kind of natural state, anterior to the civilized world, in which his hypothetical subjects lived in bestial conditions.

Condillac's conception of the origin of language was mainly focused on a generative reconstruction of the human mind, thus emphasizing the relation between language and knowledge. In contrast to Condillac, Rousseau, in his *Discours sur l'origine et les fondements de l'inégalité parmi les hommes* (1755), aimed at a generative reconstruction of human society.

According to Rousseau, the development of human society played a decisive rôle in the formation of language, but he was unable to cope with the difficult question of the anteriority of language or society, thus entangling himself in a vicious circle and leaving to others the solution of the problem (cf. Starobinski, 1971; Droixhe/Haßler, 1989; Neis, 2003).

Rousseau integrated his theory of language into his conception of a natural state of mankind, prior to the formation of civilization. For his account of the state of nature, Rousseau followed the ideas traced by the 17th-century natural law theorists such as Thomas Hobbes (1588–1679) (see Figure 6), Hugo Grotius (1583–1645), and Samuel Pufendorf (1632–1694). However, Rousseau rejected Hobbes' vision of the state of nature as a condition of permanent war by introducing his natural man (*homme de la nature*) as a creature who led a vagabond life and lived in harmony with his environment. Into his hypothetical initial state of mankind, Rousseau integrated his conception of the origin of language and made language arise as a result of the primitive needs of man. Following this Epicurean vision, Rousseau also adopted Condillac's theory of the *langage d'action*, a first step of communication based on gestures and instinctive cries (cf. Haßler, 1984).

The vision of society and language offered by Rousseau in the *Discours* had a strong impact on the discussion of the origin of language in the second half of the 18th century, especially on the prize contest on the origin of language organized by the Berlin Academy in 1769 (cf. Megill, 1974). After a long internal discussion, the Academy invited scholars from across Europe to submit their essays on the origin of language. As language was considered an essential component of the natural constitution of man, it was required to reveal how man might have invented it and whether he might have been able to create it on his own.

This question has also to be seen in the light of a discussion at the Berlin Academy which was caused by a lecture given in 1756 by Johann Peter Süßmilch (1707–1767): the *Versuch eines Beweises, daß die erste Sprache ihren Ursprung nicht vom Menschen, sondern allein vom Schöpfer erhalten habe*. In this text, Süßmilch, himself a member of the Berlin Academy, defended the divine origin of language and protested against all conceptions of the origin of language based on the Epicurean doctrine. Because of its high degree of perfection, language could only have been the work of God. Primordial man, lacking the higher dispositions of the understanding, would never have been capable of inventing a system of such a complexity as language.

By announcing the prize topic as being the origin of language in 1769, the Berlin Academy endeavored to find a persuasive solution that defended the possibility of human invention. Scholars considered the question as highly important, 31 essays were submitted to the Academy, an unusually high number of entries. Today, 24 of the manuscripts are still to be found in the archive of the *Berlin-Brandenburgische Akademie der Wissenschaften*. They constitute an important *text series* in the field of language origin study. Only the winning essay of Johann Gottfried Herder (1744–1803) (see Figure 7) has traditionally been considered a milestone of the Enlightenment's conception of language origins. According to Herder, the emergence of language was due to the so-called *Besonnenheit*, an intrinsic disposition of man predating reflection and permitting man to listen carefully to the sounds of his surroundings and to retain the most impressive sounds (e.g., the bleating of a sheep) as internal marks (*innere Merkworte*).

An analysis of the remaining entries (cf. Neis, 2003) revealed the existence of recurrent *topoi* and commonplaces in the arguments of the different participants. As the genesis of language is beyond historical



Figure 6 Hobbes. Drawn by Klaus Froese.

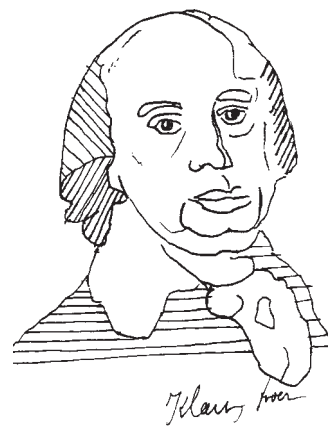


Figure 7 Herder. Drawn by Klaus Froese.

times, it can only be guessed at by means of hypothesis. Among the most recurrent *topoi* used by the participants on the origin of language question were the process of language acquisition among infants and the analysis of several hypothetical groups whose acquisition of language was impaired, such as *feral children*, the deaf and dumb, and exotic peoples. It was social, physical, or cultural deprivation that caused the exclusion of these individuals from a normal process of language acquisition.

The Origin of Language Topic in the 19th and 20th Centuries

Although the Berlin Prize contest had highlighted the debate on the origin of language in the 18th century, in the course of the 19th century, scholars continued to engage themselves in the discussion (see the contributions of Jacob Grimm (1785–1863), Wilhelm von Humboldt (1767–1835), Max Müller (1823–1900), Lazarus Geiger (1829–1870), etc.). Innovative theories were developed, partly influenced by the Darwinian paradigm and inspired by the solid empirical studies in emerging fields such as phrenology and neurology. The flood of contributions to the problem of the origin of language incited the *Société linguistique de Paris* in 1866 to ban the academic treatment of this topic within its institutional confines. By the end of the 20th century, this verdict became obsolete, and research on the origin of language was not limited to the language sciences, but became a pivotal topic for several interdisciplinary scientific domains.

See also: Aristotle and Linguistics; Conventions in Language; Plato and His Predecessors.

Bibliography

- Aarsleff H (1974). 'The tradition of Condillac: the problem of the origin of language in the eighteenth century and the debate in the Berlin Academy before Herder.' In Hymes D H (ed.) *Studies in the history of linguistics. Traditions and paradigms*. Bloomington, IN: Indiana University Press. 93–156.
- Archiv der Berlin-Brandenburgischen Akademie der Wissenschaften: Signaturen I-M 663 bis I-M 686 (Preisschriften 1771).
- Bach R (1977). 'Langue et droit politique chez Jean-Jacques Rousseau.' *Beiträge zur Romanischen Philologie* XVII Heft 1. 1977: 123–125.
- Borst A (1995). *Der Turmbau von Babel. Geschichte der Meinungen über Ursprung und Vielfalt der Sprachen und Völker*. 4 Bde. München: Deutscher Taschenbuch Verlag.
- Derathé R (1970). *Jean-Jacques Rousseau et la science politique de son temps*. Paris: Vrin.
- Derathé R (1979). *Le rationalisme de Jean-Jacques Rousseau*. (Réimpression de l'édition de Paris, 1948). Genève: Slatkine Reprints.
- Droixhe D & Haßler G (1989). 'Aspekte der Sprachursprungsproblematik in Frankreich in der zweiten Hälfte des 18. Jahrhunderts.' In Gessinger J & von Rahden W (eds.) *Theorien vom Ursprung der Sprache*. Bd I. Berlin: de Gruyter. 312–358.
- Duchet M (1971). *Anthropologie et histoire au siècle des lumières. Buffon, Voltaire, Rousseau, Helvétius, Diderot*. Paris: François Maspero.
- Duchet M & Launay M (1967). 'Synchronie et diachronie: l'Essai sur l'origine des langues et le second Discours.' *Revue internationale de philosophie* 82. Vingt et unième année. Revue trimestrielle. Fascicule 4. 421–442.
- Formigari L (1990). *L'esperienza e il segno. La filosofia del linguaggio tra Illuminismo e Restaurazione*. Roma: Editori Riuniti.
- Gensini S (1999). 'Epicureanism and naturalism in the philosophy of language from Humanism to the Enlightenment.' In Schmitter P (ed.) *Sprachtheorien der neuzeit I. Der epistemologische kontext neuzeitlicher sprach- und grammatiktheorien*. Tübingen: Narr. 44–92.
- Gessinger J (1994). *Auge & ohr. studien zur erforschung der sprache am menschen*. Berlin/New York: de Gruyter.
- Gessinger J & von Rahden W (eds.) (1989). *Theorien vom ursprung der sprache*. 2 Bde. Berlin: de Gruyter.
- Harnack A von (1901). *Geschichte der königlich-preußischen Akademie der Wissenschaften zu Berlin*. 3 Bde. Berlin: Reichsdruckerei.
- Haßler G (1997). 'Sprachtheoretische Preisfragen der Berliner Akademie in der zweiten Hälfte des 18. Jahrhunderts. Ein Kapitel der Debatte um Universalien und Relativität.' *Romanistik in Geschichte und Gegenwart, Jahrg. 3, Heft 1*. 3–26.
- Haßler G & Schmitter P (eds.) (1999). *Sprachdiskussion und Beschreibung von Sprachen im 17. und 18. Jahrhundert*. Münster: Nodus Publikationen.
- Juliard P (1970). *Philosophies of language in eighteenth-century France*. The Hague/Paris: Mouton.
- Megill A D (1974). *The enlightenment debate on the origin of language*. New York: Columbia University.
- Moravia S (1984). *Il ragazzo selvaggio dell'Aveyron. Pedagogia e psichiatria nei testi di J. Itard, Ph. Pinel e dell'anomimo della 'décade'* (Segni di Segni. Quaderni di Filosofia del linguaggio e Antropologia culturale dell'Università di Bari, 14). Bari: Adriatica editrice.
- Neis C (1999). 'Zur Sprachursprungsdebatte der Berliner Akademie (1771). Topoi und charakteristische Argumentationsstrukturen in ausgewählten Manuskripten.' In Haßler G & Schmitter P (eds.) *Sprachdiskussion und Beschreibung von Sprachen im 17. und 18. Jahrhundert*. Münster: Nodus Publikationen. 127–150.
- Neis C (2002). 'Francesco Soave e la sua posizione sull'origine del linguaggio: dal dibattito all'Accademia di Berlino (1771).' In Gensini S (ed.) *D'uomini liberamente parlanti. La cultura linguistica italiana nell'Età dei Lumi e il contesto intellettuale europeo*. Roma. Editori Riuniti. 191–218.

- Neis C (2003). *Anthropologie im Sprachdenken des 18. Jahrhunderts. Die Berliner Preisfrage nach dem Ursprung der Sprache (1771)*. Berlin: de Gruyter.
- Ricken U (1984). *Sprache, Anthropologie, Philosophie in der französischen Aufklärung*. Berlin: Akademie-Verlag.
- Schlieben-Lange B (1984). 'Vom Vergessen in der Sprachwissenschaftsgeschichte. Zu den "Ideologen" und ihrer Rezeption im 19. Jahrhundert.' *Zeitschrift für Literaturwissenschaft und Linguistik* 14(53/54), 18–36.
- Schmitter P (1987). *Das sprachliche Zeichen. Studien zur Zeichen- und Bedeutungstheorie in der griechischen Antike sowie im 19. und 20. Jahrhundert. (Studium Sprachwissenschaft. Beiheft 7.)* Münster: MAkS Publikationen.
- Schmitter P (ed.) (1991). *Sprachtheorien der abendländischen Antike. (Geschichte der Sprachtheorie 2.)* Tübingen: Narr.
- Schmitter P (ed.) (1999). *Sprachtheorien der Neuzeit I. Der epistemologische Kontext neuzeitlicher Sprach- und Grammatiktheorien*. Tübingen: Narr.
- Starobinski J (1971). *Jean-Jacques Rousseau: la transparence et l'obstacle. Suivi de sept essais sur Rousseau*. Paris: Gallimard.
- Sulek A (1989). 'The experiment of Psammetichus: fact, fiction, and model to follow.' *Journal on the History of Ideas*. Oct.–Dec. 1989. 645–651.
-

P

Paradoxes, Semantic

T Kenyon, University of Waterloo, Waterloo, ON, Canada

© 2006 Elsevier Ltd. All rights reserved.

The semantic paradoxes are a group of logico-linguistic puzzles characterized by the key roles of semantic notions such as truth, reference, or meaning in their construction. Like any paradox, their defining feature is that they force us to accept absurd consequences without employing any premise or concept that can easily be rejected. The semantic paradoxes are usually grouped together for the purpose of distinguishing them from the logical or set-theoretic paradoxes, the most famous of which are Russell's Paradox, Cantor's Paradox, and the Burali-Forti Paradox. The depth of difference between these two groups is not always clear, however, so it is unwise to attach much significance to the distinction.

The Paradoxes

The most significant of the semantic paradoxes, the Paradox of the Liar, is also the most ancient, dating at least to the 4th century B.C.E. Eubulides of Miletus is the first known writer to have formulated it as a fairly explicit logical puzzle. In a still earlier form, the Paradox (often simply called 'the Liar') consisted of an utterance attributed to the philosopher Epimenides, who was said to have framed the implicitly puzzling statement 'All Cretans are liars.' The idea was that since Epimenides himself was a Cretan, his statement, if true, would be false. An allusion to this forerunner of the puzzle even surfaced in a biblical reference from the letter to Titus, whose author seemed to take Epimenides' example rather too literally: "Even one of their own prophets has said, 'Cretans are always liars...'" In any case, this version turns out not to be very obviously paradoxical, since one need not always lie in order to count as a liar. One could be a lying rascal in general who happens to be telling the truth just this once.

Avoiding some of these quibbles, Eubulides simply asked whether someone who says 'I am lying' speaks truly or falsely. This is a better paradox, though still somewhat confounded by the fact that while lying entails speaking a falsehood, the converse does not hold. If the speaker is lying, then by the meaning of 'lying' the utterance is false; but if the utterance is false, then it does not quite follow that what he says – namely, that he is lying – is true. One can speak falsely without lying. The most obviously paradoxical version of the puzzle, then, is the still simpler case 'This sentence is false.' While this may be intuitively paradoxical, we can make the problem transparent by adding explicit assumptions: Any well-formed sentence *P* in the indicative mood is true just in case *P* and is false otherwise; the Liar sentence is such a well-formed sentence; and the expression 'this sentence' in the Liar sentence refers to the Liar sentence itself. So assume that the Liar sentence is true. What it tells us, though, is that it is false. So if it is true, then it is both true and false. Now assume that it is false. Then, since the Liar sentence tells us that it is false, what it tells us is true; if false, it is both true and false. Whether we take the Liar to be true or false in the first instance, we are forced us to accept a contradiction. Yet by our first assumption it must be either true or false, so we seem committed to the contradiction. Variants of the paradox include the Strengthened Liar ('This sentence is not true'), which dispenses with the assumption that 'false' means 'not-true,' and sets of sentences as in the postcard paradox (supposing a postcard that reads on one side 'The sentence on the other side of this card is true,' and on the other side, 'The sentence on the other side of this card is false'). Various responses to the Liar are summarized later in this article.

Other well-known semantic paradoxes have more recent origins. The Grelling-Nelson Paradox (sometimes just known as Grelling's Paradox, or the Heterological Paradox) dates from a 1908 paper discussing Russell's Paradox and the Burali-Forti Paradox. Grelling and Nelson's example is based on

a seemingly innocuous adjective. Adjectives, of course, are descriptive terms. Among the things they can describe are words, and among the words an adjective might describe is that adjective itself. So, for example, the adjective 'linguistic' is linguistic, just as the adjective 'short' is short and 'polysyllabic' is polysyllabic. Call adjectives that describe themselves *autological*. On the other hand, most adjectives do not describe themselves. 'Long' is not long, 'wet' is not wet, and 'plebian' is not plebian. Call adjectives that do not describe themselves *heterological*. Notice, though, that both 'autological' and 'heterological' are themselves adjectives. So are these two adjectives themselves autological or heterological? In particular, is 'heterological' heterological? In order to be heterological, the adjective must not describe itself and hence not be heterological; yet if not heterological, it is non-self-describing, and hence is heterological. It describes itself if and only if it does not.

A semantic paradox of definability is Berry's Paradox. There are two versions of it, the lesser known being the more historically accurate. In 1904 Berry wrote to Russell, suggesting a paradoxical definition: "the smallest ordinal number that cannot be defined in finitely many words." Of course, this definition defines such a number in finitely many words, yielding a contradiction. Russell later published a different paradox, which he named after Berry and which is the more familiar version: Consider the least integer not nameable in fewer than 19 syllables. We just defined (loosely, named) this integer in 18 syllables, however, so a contradiction is forthcoming: the least integer not nameable in fewer than 19 syllables is nameable in fewer than 19 syllables.

Another paradox of definability, resembling both Berry's and Cantor's paradoxes, is Richard's Paradox. Imagine a numbered infinite list of all English phrases that pick out exactly one positive real number. Now consider this definition: the positive real number having in its decimal expansion, at the n th position after the decimal, a 1 if the number defined by n th entry on the list has a 0 in its n th position after the decimal, and a 0 otherwise. This seems to define a number and hence must be an entry in the list. By hypothesis, however, any number thus identified cannot be defined by any entry on the list, since for any n th entry on the list, the Richard number differs from that entry in the n th position after the decimal. So this definition cannot be an entry in the list.

Both Berry's and Richard's paradoxes are usually diagnosed as arising from inherent imprecision in the concept of definition. Whether some definition is

formulable (and in how many words or syllables) is clearly contingent upon features of the particular language in question. When a suitably precise definition is given for definability itself – explicitly relativized to a language – that is, *definable in L* – the air of paradox is dispelled.

Suggested Solutions to the Liar

An analogous explanation is typically applied to the standard example of a semantic paradox, the Liar, the literature on which is both the most extensive and the most intellectually fertile of any on the various paradoxes. The most significant and technically developed interpretation of the Liar diagnoses the problem in terms of imprecision in the semantic notions found within natural language, and particularly the notion of truth.

Alfred Tarski gave the first rigorous expression of this line of thought, arguing that a natural language like English, with its apparently univocal predicate 'is true,' is a semantically closed language. This means that when one predicates truth of a declarative sentence S in English, the result is another English sentence, S is true. And this new sentence is in turn subject to the application of the same truth predicate – and the same, *mutatis mutandis*, goes for the predicate 'is false.' Tarski argued that the contradictions arising from the Liar are an artifact of semantic closure, hence showing that natural languages cannot be understood as employing a logically coherent concept of truth. The coherent idea approximating the ordinary notion is really that of *truth in L* : truth relativized to a language, where this means a formally defined language rather than a natural language. We have to refine our use of the truth predicate by viewing it as the application of a *metalanguage* to an *object language*, in the terminology Tarski coined. Of course, since any metalanguage can in turn be characterized by yet another language, the designations of object language and metalanguage are relative. On this view, to apply 'is true' to a sentence in a logically consistent way is to use a language of a semantic level distinct from that of the sentence to which the predicate is applied: we apply a metalanguage predicate to an object language sentence. Reading the Liar sentence as an attempt to apply the truth predicate of L to a sentence of L , we can use the Tarskian diagnosis to conclude that the Liar is simply ill formed rather than paradoxical.

Tarski's proposal and its ancillary details have been an important source of insights in logic and the

philosophy of language. Notwithstanding its great influence, however, it is not generally accepted as a solution. Some writers object that natural language does not seem to contain the hierarchy of notional languages defined by Tarski. It is not clear that this complaint engages Tarski's precise view, however; he claimed, rather, that the machinery of object language and metalanguage is *imposed* on fragments of natural language for analytic purposes. Another class of objections is perhaps closer to the mark: Tarski's conclusion that the truth predicate must belong to a language 'essentially richer' than the language of which it is predicated is neither perfectly clear in its formal definition nor entirely plausible as applied to many Liar-like cases.

As Saul Kripke argued, there seem to be many cases in which linguistically internal applications of a univocal truth predicate are not ill formed. Suppose Jill's only utterance about Larry is 'Most of what Larry says about me is true.' Larry, then, might say only the following about Jill: 'Jill is a banker; Jill went to Athabasca College; everything Jill says about me is false.' In the event that Larry's first two statements are true, there is no temptation to find ill-formedness in either utterance: Jill's statement is true, and Larry's third statement is straightforwardly false. But should only one of Larry's first two statements be correct, then Jill's statement will be true just in case Larry's third statement is true. Yet this says Jill's statement is false, making Larry's third statement false as well; it is now essentially the postcard version of the paradox. Here the Tarskian analysis seems misplaced. For how could nonlogical, nonlinguistic facts like Jill's profession determine whether her statement and Larry's third should count as inherently ill formed? Kripke's solution was to reject the principle of bivalence that enables the inference from *S is not true* to *S is false*. Both truth and falsity attach to grounded statements, by which Kripke meant those whose truth conditions do not ultimately implicate truth itself. Ungrounded statements, including the Liar, are neither true nor false.

Kripke's solution, too, is not regarded as decisive. At the metalinguistic level, at least, Kripke's theory seems to inherit the *Strengthened Liar* paradox: 'This sentence is not true.' Intuitively this sentence correctly describes itself whether it is false or ungrounded, and so is true just in case it is not true. The Strengthened Liar is generally taken to be the strongest version of the paradox, the version that any successful solution must address. There is currently no consensus on

how best to go about formulating such a solution. Some devotees of the puzzle continue in the Tarskian tradition, while some advocate an explanation in terms of language pragmatics or faulty presupposition; utterances of the Liar are said to amount to a sort of performative misfire on this view. Still other approaches proceed via nonstandard logics, an example of which is *dialetheism*, a kind of paraconsistency that tolerates the existence of statements that are both true and false. The Liar is suggested as an example of such a statement. As a paraconsistent approach, dialetheism rejects *ex falso quodlibet*, the classical 'explosion' principle by which the consequences of contradictions ramify without constraint. The dialetheist proposal depicts the Liar as a surprising and idiosyncratic, but logically tolerable, true-and-false statement while limiting permissible inferences from such contradictions in order to keep the idiosyncrasy localized. It is unclear what motivation can be given for this proposal, however, except for the difficulties of other attempted solutions.

See also: Logical Consequence; Metalanguage versus Object Language; Presupposition; Semantic Value; Truth; Theories of in Philosophy.

Bibliography

- Barwise J & Etchemendy J (1987). *The liar: an essay in truth and circularity*. Oxford: Oxford University Press.
- Gupta A (1982). 'Truth and paradox.' *Journal of Philosophical Logic* 11, 1–60.
- Kripke S (1975). 'Outline of a theory of truth.' *Journal of Philosophy* 72, 690–716.
- McGee V (1991). *Truth, vagueness, and paradox: an essay on the logic of truth*. Indianapolis: Hackett Publishing.
- Priest G (1984). 'The logic of paradox revisited.' *Journal of Philosophical Logic* 13, 153–179.
- Prior A (1961). 'On a family of paradoxes.' *Notre Dame Journal of Formal Logic* 2, 16–32.
- Quine W V (1966). 'The ways of paradox.' In *The ways of paradox and other essays*. New York: Random House. 3–23.
- Russell B (1908). 'Mathematical logic as based on the theory of types.' *American Journal of Mathematics* 30, 222.
- Sainsbury R M (1995). *Paradoxes*. Cambridge, UK: Cambridge University Press.
- Strawson P F (1949). 'Truth.' *Analysis* 9, 83–97.
- Tarski A (1956). 'The concept of truth in formalized languages.' In *Logic, semantics, metamathematics*. Oxford: Clarendon Press. 152–278.

Philosophy of Linguistics

P Carr, Université Paul Valéry, Montpellier, France

© 2006 Elsevier Ltd. All rights reserved.

The philosophy of linguistics concerns itself with the nature of human language and of linguistic inquiry. The central issues concern epistemology and ontology. 'Epistemology' is the branch of philosophy dealing with the nature of knowledge – in this case linguistic knowledge. 'Ontology' is the branch of philosophy dealing with reality, with what exists (is there only one kind of reality, or should we speak of a pluralism of realities?). In the philosophy of linguistics, the main ontological question is whether we can reasonably claim that there are linguistic realities, or a single linguistic reality, as distinct from other kinds of reality. Epistemology and ontology are closely intertwined, especially if we assume that there are linguistic realities, and that they constitute a kind of knowledge. The philosophy of linguistics is closely connected with issues in the philosophy of science, which concerns itself with questions such as: is there such a thing as scientific knowledge, distinct in kind from other sorts of knowledge? Is it possible to distinguish scientific theories from other sorts of theory? If so, how? Is there such a thing as a scientific method? If so, how do we characterize it? What makes it different from other methods? Clearly, we cannot say whether, or to what extent, linguistic inquiry might be scientific unless we have some sense of what we mean by 'science.' Furthermore, given that philosophers of science disagree on the answers to the questions just raised, linguists must adopt a particular philosophy of science if they are to argue that linguistic inquiry is scientific in nature.

To set out the range of positions that have been proposed in the philosophy of linguistics, let us begin with a distinction made within the philosophy of science literature and then relate it to positions adopted within linguistics. Some philosophers of science, notably Feyerabend (1975), argued that there is no valid distinction to be drawn between scientific and nonscientific knowledge, that the idea of 'scientific method' is an illusion. Those who have denied this include Popper (1959), who argued that there is such a thing as scientific method, and thus scientific knowledge, distinct from nonscientific knowledge. But he disagreed with the views of the members of the Vienna Circle, a group of scientists and philosophers working in the Vienna in the mid-20th century, who embraced a position known as 'logical positivism.' The logical positivists (e.g., Moritz Schlick) argued that scientific knowledge is firmly and exclusively based on

observation, and that, because of this, scientific theories, unlike nonscientific theories, can be proven, yielding knowledge that was certain (indubitable). They went further than this: they argued that statements that are not scientific (on their understanding of the term) are literally meaningless. Their position is often referred to as 'verificationism,' since they believed that scientific theories could be verified (proven) via observation of the physical world. Although Popper believed, like the logical positivists, that there was such a thing as a scientific method, and thus scientific knowledge, he disagreed with the idea that science yields certain knowledge via verification. Instead, he argued for 'fallibilism,' the idea that our scientific hypotheses are always fallible, and that we can never have knowledge that is absolutely certain. Popper rejected verificationism and adopted, in its place, 'falsificationism': he believed that the hallmark of scientific hypotheses was falsifiability: if a claim about the world was framed in such a way that it admitted of counterevidence, then it was falsifiable and thus scientific. There is a connection between this idea and the idea of the content of a hypothesis: the more a hypothesis excludes (the more states of affairs it rules out), the more it is claiming about the world. Popper believed that scientific inquiry rested on a process of hypothesis formation, deduction of the consequences of our hypotheses, and testing of those consequences. This is known as the 'hypothetico-deductive method.' If a falsifiable hypothesis is tested and falsified, then we may retain it until a better hypothesis emerges, abandon it, or modify it. For Popper, this was what allowed us to make scientific progress. But even if a testable hypothesis has not been falsified, that does not warrant the conclusion that it is true: for Popper, our scientific knowledge was always open to the possibility of being falsified and abandoned. Popper also rejected the view that nonscientific claims were meaningless. In its place, he argued that a given idea may begin by being nonscientific (unfalsifiable) but nonetheless meaningful, to being falsifiable, and thus scientific. An example he gave of this was the idea that the earth revolved around the sun, an idea that has its roots in Greek mythology but that eventually became scientific.

Popper's views on the scientific method were combined with another central plank in his philosophy of science: 'scientific realism.' By this he meant the idea that there are aspects of reality that are amenable to scientific inquiry but are not themselves directly observable. Examples of these are gravitational force and electromagnetic forces: we cannot, in principle, observe these, but we can observe their effects. For

Popper, it was thus valid for scientists to postulate unobservable aspects of reality and to make testable claims about them. This clearly flew in the face of the positivist claim that scientific knowledge concerns only that which is observable. Opponents of scientific-realism adopted a position referred to by Popper as 'instrumentalism,' since they argued that theoretical constructs in science were no more than that: they were instruments for systematizing and making predictions about observable phenomena. Under an instrumentalist approach, we are not warranted in saying that our constructs correspond to theory-external objects and events in the world. The term 'phenomenalism' is also used for this position. Thus, all talk of unobservable 'forces' is nothing more than a way of talking about observable phenomena: the instrumentalist withholds any ontological commitment of the sort that claims that various forces actually exist.

The distinction between positivism, with its instrumentalist, antirealist outlook, and scientific realism had an effect on modern linguistics. During the mid-20th century, many linguists (notably, the 'post-Bloomfieldians') working in the United States argued for an instrumentalist interpretation of linguistic constructs, such as 'the phoneme' and 'the morpheme.' It is arguable that they had been influenced by the logical positivists, either directly or indirectly. In insisting that linguistics, to be scientific, had to concern itself only with that which was observable, they were committed to a position that was antimentalistic, since mental states and processes are, by definition, unobservable. This outlook was combined with a positivistic conception of psychology, often referred to as 'behaviorism,' the view that a properly scientific psychology should concern itself only with observable behavior and not with unobservables.

It is arguable that much of Noam Chomsky's thinking (1965, 1993, 2002 and elsewhere) about the nature of linguistic inquiry is in line with Popper's thinking about science. Like his immediate predecessors in the United States, Chomsky believed that the kinds of linguistic inquiry he was engaged in (generative linguistics) were scientific in nature. But, unlike them, Chomsky adopted scientific realism: he argued that science is not limited to observables. It is Chomsky's scientific realism that allowed him to embrace mentalism in the study of language: freed from the insistence that science concerns itself only with that which is observable, Chomsky could allow that there are linguistic realities that are mental in nature, in stark contrast to the antimentalistic stance of his immediate predecessors. He went further than this: he insisted that linguistic realities are exclusively mental. More specifically, they are mental states, not processes, and they are strictly internal to individuals. Thus,

linguistic inquiry was a branch of individual psychology for Chomsky. This clearly flies in the face of any social conception of the object of linguistic inquiry, which Chomsky labeled 'E-language,' where the 'E' means 'external to individual minds.' In contrast, he insisted that the object of inquiry in linguistics is 'I-language,' where the 'I' stands for both 'Individual' and 'Internal' (to the mind). This is Chomsky's 'internalism.' Connected with this is Chomsky's insistence that language is not designed for use. This is not intended to mean that language is not, in fact, used for communication (no one could deny that it is); rather, communicative use is not what language is for. This is Chomsky's 'antifunctionalism': in his view, language is not driven by communicative function. Rather, language is for thinking: language and thought are intimately connected, for Chomsky. Connected with this is the long-standing distinction between 'competence' and 'performance': for Chomsky, the object of inquiry is not observable utterance phenomena (performance), or acts of online mental processing of utterances (which also falls within performance), but the knowledge (competence) that allows performance (use of that knowledge) to take place. The choice of terminology was unfortunate, since in everyday parlance, 'competence' denotes the ability to perform in some domain, as in *John's competence as a manager is unquestionable*. But Chomsky long since insisted that his conception of linguistic knowledge was not to be interpreted as knowing how to do something; nor was it to be interpreted as knowing that something is the case. Rather, linguistic knowledge was an unconscious mental state that grew in the mind of the child.

This notion of growth is a biological one: according to Chomsky, language 'acquisition' (the term is inappropriate in discussing Chomsky's views, since, for Chomsky, language is not acquired – it is innate) was not something that a child did: it was something that happened to the child. In this view, 'language acquisition' is a process of biological growth. This constitutes Chomsky's 'naturalism': the view that language belongs to the natural world and not to culture, a view that is controversial. A central component of Chomsky's naturalism was his claim that human beings are born with linguistic knowledge, that there is innate cognitive content that is specifically linguistic. This is often referred to as the 'innateness hypothesis,' although Chomsky disliked the term on the grounds that everyone agreed that there are innate cognitive capacities: the question is what they might be. In arguing for innate linguistic knowledge, Chomsky adopted a version of 'rationalism,' often referred to as 'nativism.' In the history of philosophy, rationalism is most frequently associated with the

work of René Descartes, who argued that not all human knowledge is acquired through interaction with the mind-external world. Rather, he argued that there is innate cognitive content, often referred to as the 'Cartesian doctrine of innate ideas.' Chomsky explicitly associated himself with the Cartesian tradition and has elaborated a specifically linguistic version of rationalism. The Chomskyan version of the doctrine of innate ideas is that we are born with a universal set of semantic primitives, out of which specific word meanings are constructed. Although the terms 'nativism' and 'rationalism' tend to be used interchangeably, some definitions of nativism claim only that it amounts to allowing that there are innate cognitive capacities, a view that is consistent with 'empiricism,' a doctrine that has various versions, but that emphasizes the role of our experience with the mind-external world in acquiring knowledge. Empiricists need not deny that we are born with innate cognitive capacities of various sorts (such as the capacity for forming inductive and analogical generalizations); where empiricists in linguistics disagree with Chomsky is in denying that innate, specifically linguistic, knowledge exists.

Another central plank in Chomsky's philosophy of linguistics is his adoption of 'modularism,' the view that at least some aspects of the mind are modular in nature. By 'module' he meant a distinct component of mind devoted to a specific cognitive domain (such as language, or recognition of familiar faces, or the visual system as a whole). The notion of modularity of mind is associated with the work of Fodor (1983), who suggested a set of defining properties of mental modules, including the property of being 'encapsulated.' The idea behind encapsulation is that the internal workings of a mental module are unavailable to other aspects of the mind. However, the output of a module is available to other aspects of the mind. For instance, we cannot access the inner workings of the module of mind that compels us to suffer optical illusions, but we do access the output of that module. Like Fodor, Chomsky never believed that the mind is entirely modular, but he did consistently espouse linguistic modularity: he believed that the mind contains an innately endowed language module. He also followed Fodor in believing that the mind contains a central processor that is nonmodular and that is used in, among other things, the fixation of beliefs. Our experience of optical illusions provides us with an example of the distinction between the modular visual system and the central processor: while the central processor can arrive at the belief that a given picture constitutes an optical illusion, it cannot override the output of the visual system. Knowing that a picture is illusory does not allow us to escape from

experiencing the illusion. Fodor claimed that the central processor contains a 'Language of Thought' (LoT), which is said to be innate. The LoT is said to be a universal set of semantic primitives that has a syntax of its own, allowing for the construction of complex concepts out of a set of semantic primitives. One of the arguments in favor of the LoT is the claim that language acquisition involves entertaining hypotheses about the structure of the language to be acquired: such hypotheses cannot be formulated unless there is an innate LoT for them to be formulated in. One important difference between Chomsky and Fodor is that, when Fodor spoke of language as a mental module, he conceived of it as an input/output module that receives certain kinds of output from the mind-external world, processes it, and creates an output. For Chomsky, input/output modules were distinct in kind from the innermost cognitive system, which constitutes language.

An alternative recent approach to modularity can be found in the work of Karmiloff-Smith (2001), who argued that, rather than postulating innate mental modules, we should postulate, in addition to general cognitive capacities, innate biases in certain domains, which then evolve into a modular organization. This notion of emergent modularity is part of Karmiloff-Smith's 'constructivism,' according to which a child actively constructs a mental grammar, revising his or her own internal representations during the course of development. This stands in contrast to Chomsky's passive conception of language acquisition. The work of Karmiloff-Smith occupies a midway position between rationalism and empiricism. A more overtly empiricist approach to linguistic knowledge was adopted by the linguist Geoffrey Sampson (1997), who opposed linguistic rationalism for several decades. Sampson claimed that the arguments given in support of nativism do not hold up to close scrutiny. Instead, he argued that the child starts with a blank slate, but with the capacity to learn via the hypothetico-deductive method. On this view, language acquisition is learning, a view entirely at odds with the Chomskyan view.

Others have objected to the ontological status of linguistic objects proposed by Chomsky. Among these are Katz (1981), who argued for Platonism in linguistics. Arguing against Chomsky's psychologism (the view that linguistics is a branch of psychology), Katz's view, which is a version of realism, was that there are linguistic realities but they are 'abstract,' in the sense of being Platonic in nature: they are not spatio-temporal (they do not exist in space and time). Katz was a semanticist who argued, among other things, that the existence of necessary truths in language leads us inexorably to Platonism, since necessary

truths are timeless: they predate human psychological states. Katz also argued that, since the notion 'knowledge of' is a two-place predicate, the notion 'knowledge of language' presupposes linguistic objects of which we have mind-internal knowledge. Knowledge of the structure of a sentence, Katz argued, is distinct from the structure of the sentence *per se*. Engaging in acts of intuitive grammaticality judgment entailed, Katz claimed, our gaining direct access to abstract Platonic objects, rather than gaining access to mind-internal states, as Chomsky argued. Note that Chomsky's use of the term 'knowledge' did not, in fact, appeal to a two-place predicate: linguistic knowledge was not knowledge of something, for Chomsky. Katz argued not only that linguistic objects such as sentences were Platonic objects, but that specific languages such as French and Spanish also had this status. For Katz, linguistics was on a par with mathematics, as one of the 'sciences of the intuition': just as mathematical objects are (according to Katz) mind-external Platonic objects to which we can gain intuitive access, so are sentences. A less radical version of antipsychologism was the view, first proposed by Roger Lass (1976), that linguistic objects are intersubjective in nature, specifically that they belong to the ontological category referred to by Popper (1972) as 'world three.' Popper argued for 'ontological pluralism,' the view that the world is open-ended and that new sorts of reality can emerge. For Popper, 'world one' was the world of inanimate physical objects and events. 'World two' was the world of mental states, which emerged with the emergence of life forms, particularly those with minds. 'World three' was where Popper believed scientific knowledge belonged: he regarded scientific theories as objective knowledge, intersubjective in nature. His principal argument was that our scientific theories may contain logical consequences that might, at a given stage in scientific history, remain unnoticed, later to be discovered. Popper believed that such consequences are real, even before being discovered, and that they therefore must be said to exist independently of our subjective mental states. Carr (1990) later elaborated on Lass's idea, arguing that linguistic knowledge is objective knowledge in the Popperian sense, while also arguing that generative linguistics is scientific in the Popperian sense.

A distinct version of linguistic knowledge as mutual knowledge can be found in the work of Itkonen (1978), who argued that the central notion in language is the notion of socially constituted rules or norms. Itkonen distinguished spatiotemporal events such as rocks rolling down hills, from our actions, which necessarily contain a component of intentionality. It is social actions that are central to Itkonen's

thinking: he argued that socially constituted norms (rules) formed the basis for linguistic behavior. These norms (conventions) are the object of grammatical inquiry for Itkonen, and since they are social in nature, the object of grammatical inquiry is social, not mind-internal.

A more recent approach to the question of the respective roles of nature and social conventions can be found in the work of Noel Burton-Roberts (see Burton-Roberts and Carr, 1999). Burton-Roberts adopted a version of naturalism; he advocated 'radical internalism,' the view that the only coherent conception of Chomskyan internalism is one under which the contents of innate linguistic knowledge are not in any sense internalized. 'Internalization' implies setting up mind-internal representations of events or objects that are mind-external. This, Burton-Roberts argued, is not what Chomsky intend when he advocated internalism. He agreed with Chomsky that radically internal language is 'austere' in the sense that it has no access to other aspects of mind that contain internalization of properties of the mind-external world, such as linearity (sequentiality), derived from the linear sequencing of the speech signal. Knowledge of the conventions regarding the sequences of phonemes, morphemes, and words in a specific language is 'internalized knowledge' for Burton-Roberts. The place of phonology was a central concern for Burton-Roberts (see Burton-Roberts 2000), since sequencing is a defining feature of phonology. The relationship between phonology and semantics was equally important for Burton-Roberts, since that relationship was both arbitrary and conventional. He argued that if one allows, as Chomsky did, that phonological knowledge is part of radically internal linguistic knowledge, then one compromises radical internalism by incorporating that which is conventional into that which is natural. The distinction between that which is radically mind-internal and that which is internalized is crucial here. That which is radically internal to the mind has not been internalized from the mind-external world. That which is internalized, such as the mental image of a familiar face, is not radically internal: it results from internalizing aspects of the mind-external world.

Burton-Roberts distinguished a generic conception of language from the Chomskyan naturalistic conception of language. The former appealed to a notion of the universality of the notion 'language' by generalizing over all human languages: 'language' here is the generic term for the set of all languages. In this view, a specific language is an 'instantiation,' or 'token,' of the type 'language.' The generic conception of language lends itself to an instrumentalist interpretation: the notion 'language' can be seen as a construct

that has no reality beyond the specific languages which it ranges over. The naturalistic conception of language appeals to a notion of 'language' as something quite independent of particular languages: the reality of language, in this view, does not reside in particular languages. Rather, 'language' is the innate linguistic knowledge shared by all members of the species. The naturalistic conception of language lends itself to a realist interpretation: 'language' as a biological reality. The generic conception of 'language' embodies a methodological claim: that the way to understand the notion 'language' is to investigate particular languages. In this view, strong universals are properties attested in every human language that one will ever encounter, whereas weak universals are tendencies within the world's languages. The naturalistic conception, on the other hand, embodies an ontological claim: that the notion 'language' denotes an object in the biological world. Universals here are the properties of that object. Burton-Roberts argued that Chomsky's conception of I-language is inconsistent in that it vacillates between the generic and the naturalistic conceptions of language. This is so because Chomsky takes phonology, and thus linearity, to constitute part of naturalistic I-language. For Burton-Roberts, linearity and conventionality necessarily lay outside of language conceived of in naturalistic terms (since conventions are social in nature: they belong to culture, not nature; conventions are not given by biology). Crucially, Burton-Roberts argued that Chomsky was mistaken in regarding the relationship between 'language' (understood in the naturalistic sense) and particular languages as one of instantiation, since this amounted to incorporating an aspect of the generic conception into the naturalistic conception. Burton-Roberts argued that, if we are to sustain the naturalistic conception, we must replace the relation of instantiation with one of 'physical representation': specific utterance phenomena are mind-external physical representations of the contents of the language module. Burton-Roberts was at pains to point out that physical representation is entirely distinct from instantiation: that which is a physical representation of something is not thereby an instantiation of it, just as Magritte's painting of a pipe is not itself an instance of a pipe. For Burton-Roberts, particular languages are acquired, phonologically constituted, conventional systems for physically representing a single, natural, innate language. Burton-Roberts's main epistemological point was that knowledge of the representational conventions of a specific language is quite distinct from innate linguistic knowledge: it is acquired, internalized from the mind-external world. It is conventional, not natural, in character. Burton-Roberts was thus able to

accommodate many aspects of empiricism (such as implicit and explicit learning, analogical and inductive generalization) into his conception of the acquisition of a language, while retaining Chomskyan rationalism. Interestingly, Burton-Roberts's position raises the question of whether there is any distinction to be made between the idea of an LoT and the idea of an innate language module: both have been said to contain a universal set of semantic primitives and a recursive combinatorial syntax, yielding linguistic expressions that have hierarchical structure. If no distinction is to be drawn between the two, then innate language is simply the innate LoT, which is nonmodular in nature. One then ends up with a version of naturalism that denies that there is a language module, while allowing that our knowledge of conventions of physical representation may become modular in the course of development.

Another recent approach to the language-versus-culture debate surrounding the nature of linguistic knowledge is that adopted by Levinson (2001, 2003a, 2003b). Levinson and colleagues resuscitated a version of the Sapir-Whorf Hypothesis, according to which the categories of the language one has acquired shape, to some extent, cognitive content and mental activity. Levinson claimed that, on the basis of a wide sample of languages, it is possible to conclude that there are three main kinds of Frame of Reference (henceforth FoR) for expressing the spatial location of an object relative to other objects on a horizontal plane. First, there is an 'egocentric' (viewpoint-dependent) or 'relative' FoR (what counts as left or right is relative to one's location: turn around 180 degrees and right becomes left), as in *The ball is to the left of the tree*. Second, there is a 'geocentric' or 'absolute' FoR (what counts as north or south is not relative to one's location), as in *The ball is to the north of the tree*. Third, there is an 'object-centered' or 'intrinsic' FoR. This latter kind of FoR involves locating an object A with reference to the partitioning of another object B into parts and the projection of an axis from the centre of object B through one of its parts, as in *The ball is at the front of the truck* (Levinson, 2003a: 8), where object A is the ball, object B is the truck, and the relevant part of object B is the front of the truck.

Levinson claimed that it has been assumed by nativists that concepts such as 'left' and 'right' are universal, since they are necessarily innate in the nativist view. Crucial to Levinson's argument is the claim that not all languages possess all three systems, but that any given language will possess at least one. Note that the conception of universalism appealed to here is the generic conception, not the naturalistic one: Levinson's argument here does not necessarily undermine a consistently naturalistic conception of

the notion 'language.' Importantly for Levinson, some languages (e.g., Oxchuc Tzeltal, an American Indian language spoken in Mexico) have no 'relative' spatial expressions at all. In such languages, one cannot say *The ball is to the left of the tree*; rather, one has to express the spatial location by means of absolute terms, such as *The ball is north of the tree, Please pass the plate to your east*, or *Take the first turning to the south* (but not *Take the first turning on the right*). Equally, in work with Majid *et al.* (2003), Levinson and colleagues argued that speakers of the Australian language Guguyimidjir entirely lack both relative and intrinsic FoRs and use only absolute FoRs, so that even the description of the location of an object on a body part is described, and conceived of, in absolute terms, as in *There's an ant on your south leg*. Levinson argued that the conceptual representations we use in nonlinguistic thinking vary depending on the language we have acquired, a strikingly Whorfian claim that has incited opposition among broadly Chomskyan linguists such as Li and Gleitman (2002). Levinson and colleagues carried out experiments in which a set of languages were classified as absolute, relative, or mixed and required speakers of those languages to carry out nonlinguistic tasks involving the spatial location of objects. In one such task, reported in Levinson (2001: 578–579), subjects were shown three different toy animals in a row on a table and asked to memorize the order in which they appeared. The subjects were then turned around by 180 degrees, taken to another table, and asked to place the objects in the same order they had seen them in. When he compared speakers of Tzeltal, who have a language with an absolute FoR to describe objects in a row, he found that they placed the objects in the opposite order from the way they were placed by speakers of Dutch, which uses a relative FoR: the subjects seemed to be memorizing the rows of objects in a way that was consistent with the FoR given by their native language.

The claims made by Levinson are clearly of considerable importance in debates about the nature of language, since they rest on the claim that both nature and culture play a role in linguistic knowledge. For Levinson, there is coevolution of language, nature, and culture, and cultural variation subsumes linguistic variation, which results in variation in 'cognitive style.' Levinson's work offers new empirical observations on the relative roles played by nature and culture in human language.

See also: Behaviorism: Varieties; Cognitive Science and Philosophy of Language; Language as an Object of Study; Linguistics as a Science; Modularity; Philosophy of Science and Linguistics; Plato and His Predecessors.

Bibliography

- Bowerman M & Choi S (2003). 'Space under construction: language-specific spatial categorization in first language acquisition.' In Gentner & Goldin-Meadow (eds.). 387–427.
- Bowerman M & Levinson S C (2001). *Language acquisition and conceptual development*. Cambridge: Cambridge University Press.
- Burton-Roberts N (2000). 'Where and what is phonology?' In Burton-Roberts N, Carr P & Docherty G J (eds.) *Phonological knowledge: conceptual and empirical issues*. Oxford: Oxford University Press. 39–66.
- Burton-Roberts N & Carr P (1999). 'On speech and natural language.' *Language Sciences* 21, 371–406.
- Carr P (1990). *Linguistic realities*. Cambridge: Cambridge University Press.
- Chomsky N (1965). *Aspects of the theory of syntax*. Massachusetts: MIT Press.
- Chomsky N (1993). *Language and thought*. London: Moyer Bell.
- Chomsky N (2002). *On nature and language*. Cambridge: Cambridge University Press.
- Feyerabend P (1975). *Against method*. London: Verso.
- Fodor J (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Gentner D & Goldin-Meadow S (eds.) *Language and mind: advances in the study of language and cognition*. Cambridge, MA: MIT Press.
- Itkonen E (1978). *Grammatical theory and metascience*. Amsterdam: Benjamins.
- Karmiloff-Smith (2001). 'Development itself is the key to understanding developmental disorders.' In Tomasello M & Bates E (eds.) *Language development: the essential readings*. Oxford: Blackwell.
- Katz J J (1981). *Language and other abstract objects*. Oxford: Blackwell.
- Lass R (1976). 'On defining pseudo-features: some characteristic arguments for "tenseness."' In Lass R (ed.) *English phonology and phonological theory*. Cambridge: Cambridge University Press.
- Levinson S (2001). 'Covariation between spatial language and cognition, and its implications for language learning.' In Bowerman & Levinson (eds.). 566–588.
- Levinson S (2003a). 'Language and mind: let's get the issues straight!' In Gentner & Goldin-Meadow (eds.). 25–46.
- Levinson S (2003b). 'Spatial language.' In Nadel L (ed.) *Encyclopedia of cognitive sciences 4*. London: Nature Publishing Group. 131–137.
- Li P & Gleitman L (2002). 'Turning the tables: language and spatial reasoning.' *Cognition* 83, 265–294.
- Majid A, Bowerman M, Kita S, Huan D B M & Levinson S (2003). 'Can language restructure cognition? The case for space.' *Trends in Cognitive Science* 8(3), 108–114.
- Popper K R (1959). *The logic of scientific discovery*. London: Hutchinson.
- Popper K R (1972). *Objective knowledge*. Oxford: Clarendon Press.
- Sampson G (1997). *Educating Eve*. London: Cassell.

Philosophy of Science and Linguistics

C Wearing, University of Western Ontario, London, Canada

© 2006 Elsevier Ltd. All rights reserved.

If linguistics is the science of language, we might begin by asking what kind of science it is. One answer, which has been given its major impetus by the work of Noam Chomsky, is that linguistics is a branch of psychology. A second answer, in some tension with the first, is that linguistics is a branch of mathematics. There are other possibilities, too: some view linguistics as a social science like anthropology, whereas others reject the idea that it is a science at all, taking it to be continuous with literary theory. In this article, we focus on the first two answers, as these answers have both arisen within a tradition of taking language to be a generative phenomenon (something that is built up from various component parts via rules of combination). The other possibilities, by contrast, tend to view language as considerably less systematic, amenable chiefly to the description of comparatively superficial regularities.

Depending on whether one regards linguistics as psychology, then, or as mathematics, the following questions receive very different answers: What are the characteristic methods of study in linguistics and what evidence does it employ? What does linguistics aim to discover and explain? And how is linguistics related to other scientific domains? In what follows, we examine the answers that the psychological and the mathematical positions offer to these questions. But to begin, we must examine the object of that enquiry – language – and consider what a language might be, that its study should constitute psychology or, alternatively, mathematics.

The starting point of any scientific enquiry must be its object of study, because the way in which this object is conceived determines what one aims to explain. In the case of linguistics, Chomsky's work (e.g., Chomsky, 1957, 1965) brought about a substantial shift in the conception of language, importantly altering the goals of linguistic enquiry. During the first half of the 20th century, the predominant view conceived of language as an external object, the totality of marks and noises produced by language speakers. For structuralist linguists such as Ferdinand de Saussure and Leonard Bloomfield, linguistics was chiefly a descriptive enquiry, looking for patterns exhibited across the elements of the linguistic corpus. Chomsky's proposal, by contrast, was to regard language as essentially something internal, namely, as a competence underlying the speaker/

hearer's ability to produce and understand the sounds and marks on which the structuralist was focused. On this view, language is an internally represented (implicit) system of knowledge that an individual speaker possesses, rather than the collection of sounds or marks that she is able to understand. Her language – this knowledge of linguistic rules and structures – is what allows her to understand such (external) objects.

In contrast with this position, several philosophers (Lewis, 1975; Katz, 1981; Soames, 1985; Devitt, 2003) have argued that this sort of internalist view of language is mistaken – that the structuralists were not wrong to view language as external – at least when it comes to doing linguistics. Unlike the structuralists, however, they treat languages as abstract external objects, namely, collections of sentences akin to the languages of formal logic. From this point of view, Urdu and Hungarian are on a par with the language of first-order logic: they are abstract formal systems governed by recursive grammatical rules and possessing canonical interpretations that assign meanings to their sentences. But regarded as such, languages are purely mathematical objects; psychology is not relevant to describing the properties of such entities.

For the proponent of the mathematical view, then, linguistics is an investigation of the grammatical and other structural features of languages, conceived of as abstract collections of sentences. It is, of course, compatible with this view that speakers know a grammar that allows them to understand and use the language they speak, and psychology (specifically, psycholinguistics) may tell us about the linguistic abilities of speakers and hearers. But on this view, psycholinguistics and linguistics are distinct enterprises: the former studies human psychology, whereas the latter studies abstract mathematical entities that are importantly independent of human psychology.

Thus, the goal of the mathematical linguist is to exhibit a grammar for the language being studied. As a matter of mathematical fact, there will be indefinitely many equivalent grammars for any given language, so to the extent that a language is considered in abstraction from any psychological basis it might have, there is no particular requirement that the linguist seek the grammar that the speakers of the language actually know. In a similar way, there is no reason to seek a grammar that is sensitive to regularities exhibited across multiple languages. Any grammar that is descriptively adequate, i.e., that generates all and only the grammatical sentences of the language in question, satisfies the goal of the mathematical linguist.

The principal evidence available for this undertaking consists in the judgments of native speakers about

the acceptability of sample sentences: are the sentences well formed? Are they meaningful or senseless? To borrow some well-known examples, English speakers judge that sentence (1) is an acceptable sentence of their language but (2) is not. Sentences (3) and (4) are both structurally (grammatically) acceptable, but they have notable meaning-related features. Sentence (3) has two distinct interpretations, whereas (4) arguably lacks any coherent interpretation:

- (1) The cat scratched the child.
- (2) *Cat the child the scratched.
- (3) Everyone loves someone.
- (4) ? Colorless green ideas sleep furiously.

Such intuitions about the well-formedness and meaningfulness of sentences allow the linguist both to identify the class of sentences that belong to the language, i.e., the set of sentences for which a grammar must be found, and to pinpoint the grammatical and semantic features of those sentences.

Like the mathematical view, the psychological view of language takes native speaker intuitions as a central source of evidence. However, what these intuitions are evidence for is conceived somewhat differently. Recall that the Chomskyan linguist's goal is to uncover the rules and structures that constitute a speaker/hearer's actual knowledge of language. So, like the mathematical linguist, the Chomskyan linguist aims to discover a grammar for a given language, but this grammar must meet an additional constraint: it is supposed to be the grammar that speakers of the language actually know. The Chomskyan linguist seeks a grammar that is not merely **descriptively** adequate, but also **explanatorily** adequate, i.e., a grammar that speakers of the language could plausibly have learned, given their innate psychological endowment and the variety of their experiences.

To achieve the demand of explanatory adequacy, the Chomskyan linguist adopts two specific hypotheses about linguistic competence. First, a language-specific mental faculty is postulated, consisting of mechanisms that are dedicated to language production and comprehension. (Contrast this hypothesis with the view that language production and comprehension operate via the same mechanisms used for other cognitive tasks, such as contemplating a work of art or planning a trip.) The second hypothesis is that the language faculty of each child begins in a default state that contains a considerable amount of innate knowledge (called 'Universal Grammar'). (Contrast this hypothesis with the view that the child approaches language learning as a 'blank slate'.) From this state, the child is capable of learning any natural language and is 'trained' by experience to achieve a steady state that allows him or her to speak a specific language. This steady state is the mature speaker's linguistic competence.

With these hypotheses in place, the Chomskyan linguist's task is to describe both the starting state of the language faculty (i.e., the Universal Grammar) and the stable states of the language faculties of mature speakers of the various human languages. To this end, speaker/hearer intuitions constitute a vital source of evidence. Thinking about why specific constructions are not permissible has proved particularly revealing of the contours of linguistic constraints. Furthermore, cross-linguistic evidence has been invaluable for showing how patterns that look *ad hoc* in one language reflect regularities that hold across multiple languages.

At the same time, evidence going well beyond speaker/hearer intuitions has proven useful. Evidence about how language is acquired (when and in what order children use various constructions, as well as about the constructions children do not make) helps to clarify what sorts of rules children might be coming to know about their language. For example, consider the formation of questions. A rule that relies on word order, such as 'move the first verb to the front of the sentence', will not work (it's fine for 'you are happy', but it won't work for 'the woman who is walking her dog is happy'). Instead, the child must learn a rule that manipulates the underlying grammatical structure of the sentence in order to form questions correctly. And, in fact, it is observed that children do not make the errors that would suggest they are ever following the incorrect (word order-based) rule.

In a similar way, experimental work on ordinary language use and clinical investigations into linguistic pathologies (impairments to the ability to use or understand language that result from illness or injury) are becoming increasingly helpful for testing linguistic hypotheses. For example, recent work tracking hearers' eye movements has sought to illuminate such topics as the fixing of the referents of pronouns (e.g., Runner *et al.*, 2003). The idea here is that knowing where hearers look – which characters draw their eye while hearing sentences about them – provides a new window onto the syntactic hypotheses (in this case, the Binding Theory) about what pronoun referents can and cannot be. The goal is to flesh out the information gleaned from native speaker intuitions, bringing linguistic theories more closely in line with psychological hypotheses about mental processing.

The full integration of linguistics with psychology still lies some distance in the future. Nonetheless, the Chomskyan approach already imposes substantial constraints on more general psychological explanations. In particular, specifically language-dedicated processes must constitute part of the overall psychological account of the mind and the innate capacities attributed to the language faculty must eventually find explanation within the domain of evolutionary biology.

At the same time, of course, linguistic theories are themselves responsible to the constraints imposed by psychology and related fields. In principle, evidence from any quarter might prove relevant, although psychology, computational modeling, and neuroscience seem the most likely resources.

Where does this leave the mathematical and psychological views of linguistics? As mentioned above, they are certainly 'compatible' as long as one maintains that the latter is not really linguistics. From such a perspective, the constraints and goals of the psychological view are simply irrelevant to the pursuit of linguistics proper. However, there is a difficulty with maintaining such a strict separation of linguistics and psychology; namely, it requires that we be able to distinguish between linguistic facts and psychological facts. If linguistics is quite distinct from psychology, then it must be possible to delineate the mathematical linguist's abstract collection of sentences in a principled way that does not depend on psychological facts. But it is not clear that this is possible. Insofar as the mathematical linguist relies on speaker/hearer intuitions to decide which sentences belong to the language, there is an irreducible element of idealization involved in determining which intuitions count as data. Moreover, there are a range of other facts about the use of language (certain expressions are polite, whereas others are not; certain hand gestures have conventional meanings; poetry often employs words in unusual ways) and it is not clear what principles might allow the mathematical view to identify which (if any) of these facts are among the linguistic ones (see Fodor, 1981 and Antony, 2003 for discussion). As a result, the mathematical view risks relying on mere stipulation, whereas the psychological view is able to identify the linguistic facts through empirical investigation as those facts that are explained by the functioning of the language faculty.

This sharp distinction between the mathematical and psychological views is not mandatory, however. One can individuate some of the mathematical view's languages (i.e., its abstract sets of sentences) precisely by reference to the linguistic competence of (idealized) native speakers in specific linguistic communities: a certain abstract language *L* is the language spoken by speaker *S* in virtue of *S*'s linguistic competence. In this way, the powerful logical and model-theoretic tools of the mathematical linguist can be brought to bear on language *L* in some independence from the considerations of the psychological view. Difficult questions remain about the extent to which a grammar produced in this way needs to cohere with results from psychological and other

empirical research. But viewing the mathematical and psychological views as connected to this extent allows the science of linguistics to proceed in multiple directions at once, unconstrained by *a priori* restrictions on either its methods or its evidence.

See also: Data and Evidence; E-Language versus I-Language; Language as an Object of Study; Linguistics as a Science.

Bibliography

- Antony L (2003). 'Rabbit-pots and supernova: On the relevance of psychological data to linguistic theory.' In Barber A (ed.) *Epistemology of language*. Oxford: Oxford University Press. 47–68.
- Antony L & Hornstein N (2003). *Chomsky and his critics*. Oxford: Blackwell.
- Barber A (2003). *Epistemology of language*. Oxford: Oxford University Press.
- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton and Co.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky N (1986). *Knowledge of language: Its nature, origin, and use*. Westport, CT: Praeger Publishers.
- Chomsky N (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Devitt M (2003). 'Linguistics is not psychology.' In Barber A (ed.) *Epistemology of language*. Oxford: Oxford University Press. 107–139.
- Devitt M & Sterelny K (1989). 'Linguistics: What's wrong with "the right view."' *Philosophical Perspectives* 3, 497–531.
- Fodor J A (1981). 'Some notes on what linguistics is about.' In Block N (ed.) *Readings in the philosophy of psychology*, vol. II. Cambridge, MA: Harvard University Press. 197–207.
- Katz J (1981). *Language and other abstract objects*. Totowa, NJ: Rowman and Littlefield.
- Laurence S (2003). 'Is linguistics a branch of psychology?' In Barber A (ed.) *Epistemology of language*. Oxford: Oxford University Press. 69–106.
- Laurence S & Margolis E (2001). 'The poverty of the stimulus argument.' *British Journal for the Philosophy of Science* 52, 217–276.
- Lewis D (1975). 'Languages and language.' In Lewis D (ed.) *Philosophical papers*, vol. I. Oxford: Oxford University Press. 163–188.
- Runner J T, Sussman R S & Tanenhaus M K (2003). 'Assignment of reference to reflexives and pronouns in picture noun phrases: Evidence from eye movements.' *Cognition* 81(1), B1–B13.
- Soames S (1985). 'Semantics and psychology.' In Katz J (ed.) *The philosophy of linguistics*. Oxford: Oxford University Press. 204–226.

Plato and His Predecessors

R Bett, Johns Hopkins University, Baltimore, MD, USA

© 2006 Elsevier Ltd. All rights reserved.

Predecessors

The first attempts at explicit theorizing about language in ancient Greece occurred among the itinerant teachers known as the Sophists (late 5th century B.C.E.). Ideas about language are, however, at least implicit in some texts prior to this. Parmenides (late 6th–mid-5th century B.C.E.) famously declared the impossibility of speaking or thinking of what is not, which seems to presuppose a view of meaning as consisting in, or at least requiring, reference to something in the world. And a generation or so earlier Heraclitus drew attention to the seeming paradox that one of the words for a bow, a deadly weapon, was *bios*, which is identical in spelling with the Greek word for ‘life’ (DK 22B48). This remark appears to trade on the notion that one would expect a certain fitness between words and their objects. And a similar notion is apparent in the etymologizing that occurs periodically in Homer and Hesiod; for instance, Odysseus was named by his grandfather, who was angry (*odussamenos*) towards many (*Od.* 19.407), and also himself suffers the effects of the gods’ anger (*Od.* 1.62).

It is, though, in the Sophistic period that language first becomes the object of sustained reflection. Several texts from this period address the question of the origins of language, sometimes in the context of a broader account of the origins of society. A consistent picture emerging from these texts is that language was devised by humans themselves, in response to the needs of early communities. It came into being through a set of decisions about what names should apply to what objects, and in that sense is conventional rather than natural. These decisions were not necessarily seen as entirely arbitrary; the Sophist Protagoras, in Plato’s dialogue of that name (322a6), speaks of language as the product of a *technê*, ‘skill,’ analogous to building or agriculture – which implies that one set of names may be better crafted (whatever that means) than another. But still, according to this view, there were clearly many different possible and acceptable sets of names. The picture is hardly sophisticated. However, it is the first attempt in Greek thought to understand, in something like the spirit of a social scientist – as opposed, say, to appealing to divine fiat – how language could have come about.

The question of constraints on the naming of objects was further pursued by the Sophists in a field of inquiry known as *orthotês onomatôn*, ‘correctness of names.’ It is hard to reconstruct

precisely what this involved, but it seems to have been concerned, in some sense, with making the fit between words and the world as good as possible; it thus shares something with the pretheoretical concerns implied by Homer and Heraclitus. Two Sophists, Protagoras and Prodicus, were especially associated with this subject. Protagoras is said to have distinguished the grammatical genders of words, and to have been willing to criticize existing genders of words as not appropriate to the nature of their objects; *pêlêx* ‘helmet,’ for example, should be masculine, not feminine. He also drew distinctions among what would now be called speech acts – entreaty, question, answer, and command – and was again ready to correct people’s choices among these modes of discourse. Prodicus, on the other hand, was especially concerned with precise definitions of words, including precise distinctions among nearly synonymous words; reports (and parodies) of this are common in Plato and elsewhere.

Despite Protagoras’s anticipation of the notion of speech acts, the overriding conception of language in this period was as consisting purely of names applying to objects. And this led to some extraordinary claims about language. Several Sophists and their contemporaries are reported to have argued that falsehood is impossible, that there is no such thing as contradiction, or both. The essential point of these arguments is that there are only two possible options: to speak the truth or not to speak at all. The connection with the naming conception of language is clear. If all language does is name, then truth can only consist of successfully naming an object or state of affairs. But in that case, there is no prospect of successfully *referring* to something, yet saying something false *about* it, or of two people successfully referring to the same object, yet saying contradictory things about it. The only alternative to successful naming – that is, truth – on this conception is unsuccessful naming; but if naming is all language does, unsuccessful naming is not genuine language at all.

One other text from this era deserves mention. The Sophist Gorgias wrote a treatise entitled *On What Is Not* (*Peri tou mê ontos*), in which, again, extraordinary conclusions were argued for, namely: (1) there is nothing, (2) even if there were anything, it could not be known about, and (3) even if it could be known about, this knowledge could not be communicated to anyone else. There is much dispute about the purpose of this production, although Parmenides is clearly in some way the target. But the third part, on the impossibility of communication, is of particular interest for the study of language. The

main arguments in this part were, first, that words are wholly different types of entity from the objects that they name, and second, that the contents of any two people's minds are necessarily different. We need not suppose that Gorgias actually *believed* communication was impossible. However, it is clear that he was raising (without attempting to answer) the deep question of how words manage to refer to objects at all.

Plato

The Sophists' ideas on language form the context for the reflections on this subject by Plato (mid-420s–347 B.C.E.). In a number of dialogues, widely (though by no means universally) regarded as from early in Plato's career, Socrates and his interlocutors are shown trying to answer questions of the form "What is F?", where F stands for some significant ethical characteristic such as courage, piety, or virtue in general. This enterprise is not fundamentally linguistic; even though it is often described as a search for definitions, Socrates is not interested primarily in the meanings of terms, but in the real natures of the items that those terms refer to. Nonetheless, the search is understood to require sensitivity to how the terms in question are actually used, and in this respect it seems to owe something to Prodicus's interest in precise distinctions of meaning. Socrates in these dialogues occasionally professes allegiance to Prodicus; the professions are never entirely serious, but they are not without some basis in his actual procedure.

Plato takes up the issue of correctness of names in his dialogue *Cratylus*. Here Socrates is made to examine two opposing (and extreme) views. One view, that of Hermogenes, is that linguistic correctness is purely a matter of convention. Indeed, Hermogenes even ignores the connotations of society-wide agreement normally present in the word 'convention' (*nomos*), suggesting that each individual is free to use words in whatever manner he or she decides. Against this, Socrates argues that if one accepts a view of reality as fixed – to which Hermogenes readily assents – one cannot regard correctness in language as a matter of merely arbitrary choice. Cratylus, on the other hand, holds the view that some names are *naturally* correct and some are not; the correct ones fit the nature of the things they refer to, while the incorrect ones do not. A trivial example is that 'Hermogenes' is not Hermogenes' true name, because he is not in fact the son of the god Hermes. But it is of course a much more difficult question what, in general, the natural fitness of names might consist in. The view explored by Socrates and Cratylus is that most names can be analyzed etymologically into a small number of 'primary names,' and that these primary names are correct in virtue of a natural resemblance

between their sounds and their objects. It is not clear how far Plato means us to be attracted by this theory. Socrates finds much to admire in it. But he argues that a role for convention in language cannot be altogether excluded. He is also dissatisfied with the fact that, in the version of the theory promoted by Cratylus, the originators of these names were Heracliteans who held that everything was in a state of constant change – a view that he, Socrates, cannot accept. Whatever we are ultimately supposed to think about the theory, however, the examination of the purported etymologies occupies a large proportion of the dialogue, suggesting that Plato thinks it deserves very serious consideration. Given the history of etymologizing mentioned earlier – a history to which the discussion seems sometimes to allude – this is perhaps not surprising, however frustrating and alien this portion of the dialogue may seem to us.

Another significant point that Socrates makes in *Cratylus* is that an understanding of things is more important than, and indeed indispensable for, the optimal naming of those things. And, as *Cratylus* and other dialogues make clear, the most important things, in the mature Platonic conception, are the unchanging, purely intelligible Forms – entities such as Beauty itself or Goodness itself, as opposed to any of the particular items we might call beautiful or good. Now these Forms are not, of course, linguistic entities. But it is plausible to suppose that one of the reasons Plato had for believing in Forms was a linguistic reason. Given the difficulty of the Socratic search for answers to "What is F?" questions about the virtues, it is natural to wonder how we manage to understand and use terms such as 'piety,' 'courage,' or 'virtue' at all. Even though we seem to have no difficulty employing these terms in ordinary discourse, Socrates's attempts to pin down what exactly they refer to consistently end in failure. It might well have seemed an attractive solution to this puzzle to suppose that our understanding of such terms derives from our (no doubt incomplete) grasp of the corresponding Forms, which are only imperfectly exemplified in the world around us; if Forms are what the terms really refer to, then the Socratic search, focused as it is on ordinary instances, is bound to fail. Of course, the project of understanding the Forms is hardly less ambitious, as Plato readily concedes.

Plato also tackles the question of how falsehood (along with contradiction) is possible. And here he makes what should be considered a definitive breakthrough. In the *Theaetetus* the characters are shown struggling unsuccessfully with the question of how false judgment is possible. Socrates also proposes a model of language that involves names being 'woven together' into an 'account' (*logos*,

202b), and this seems to promise an advance beyond the Sophistic conception of language as doing nothing but naming. But the promise is not fulfilled; no linguistic items are recognized other than names, and the implications of structure present in the notion of a 'weaving together' (*sumplokê*) are not followed through. In the *Sophist*, however, Plato has the central character (a visitor from Parmenides's hometown, Elea) draw a crucial distinction between a name (*onoma*) and a 'thing said' (*rêma*): the name identifies the thing being spoken of, and the 'thing said' delivers information about it. Once this distinction is in place, it is easy to see how both true and false statements may be made about the same objects. The dialogue involves much broader discussion about the legitimacy of thinking or speaking of what is not; and the term 'speaking of what is not' regularly refers, among other things, to the speaking of falsehoods. But once the possibility of this is established (against Parmenides) in general terms, the specifically linguistic details of how falsehood is possible are established relatively quickly. The importance of moving beyond the naming conception of language can hardly be overestimated; until it was clearly understood that language does more than just naming, the study of grammar, syntax, or anything else to do with the structure of language was out of the question.

Bibliography

- Ackrill J L (1999). 'Language and reality in Plato's *Cratylus*.' In Fine G (ed.) *Plato 1: metaphysics and epistemology*. Oxford: Oxford University Press. 125–142.
- Barnes J (1982). *The Presocratic philosophers*. (rev. edn.). London: Routledge & Kegan Paul.
- Bostock D (1994). 'Plato on understanding language.' In Everson (ed.). 10–27.
- Broadie S (2003). 'The sophists and Socrates.' In Sedley D (ed.) *The Cambridge companion to Greek and Roman philosophy*. Cambridge: Cambridge University Press. 73–97.
- Denyer N (1991). *Language, thought and falsehood in ancient Greek philosophy*. London: Routledge.
- Everson S (ed.) (1994). *Language*. Cambridge: Cambridge University Press.
- Frede M (1992). 'Plato's *Sophist* on false statements.' In Kraut R (ed.) *The Cambridge companion to Plato*. Cambridge: Cambridge University Press. 397–424.
- Dillon J & Gergel T (trans.). (2003). *The Greek Sophists*. London: Penguin.
- Kerferd G B (1981). *The Sophistic movement*. Cambridge: Cambridge University Press.
- Mourelatos A (1987). 'Gorgias on the function of language.' *Philosophical Topics* 15, 135–170.
- Plato (1993). *Sophist*. White N P (trans.). Indianapolis: Hackett.
- Plato (1998). *Cratylus*. Reeve C D C (trans.). Indianapolis: Hackett.
- Sedley D (2003). *Plato's Cratylus*. Cambridge: Cambridge University Press.
- Sluiter I (1997). 'The Greek tradition.' In Van Bakkum W, Houben J, Sluiter I & Versteegh K (eds.) *The emergence of semantics in four linguistic traditions: Hebrew, Sanskrit, Greek, Arabic*. Amsterdam/Philadelphia: John Benjamins. 147–224.
- Sprague R K (ed.) (2001). *The older Sophists*. Indianapolis: Hackett.
- Williams B (1994). 'Cratylus' theory of names and its refutation.' In Everson (ed.). 28–36.

Plato's *Cratylus* and Its Legacy

J E Joseph, University of Edinburgh, Edinburgh, UK

© 2006 Elsevier Ltd. All rights reserved.

Plato's *Cratylus* is the first complete surviving philosophical inquiry into the arbitrariness of language, and arguably the greatest. Although it was long thought to have been an early work of Plato's on account of being 'merely' about language, heightened appreciation of its importance over recent decades has prompted scholars to redate it to his 'great' middle period. It takes the form of a debate among three of Plato's teachers, Cratylus, Hermogenes, and Socrates, about the correctness of words (the subject taught by Cratylus). Cratylus and Hermogenes are

arguing the question within the Sophists' staple pedagogical dichotomy of *physis* 'nature' versus *nomos* 'convention.' Socrates joins them, and they invite him to adjudicate as to who is right: Cratylus, who holds that a word is correct only if naturally connected to its meaning, or Hermogenes, who thinks that any word can designate anything just as well as any other.

They begin by considering proper names, which seem obviously conventional since chosen willfully for individuals, usually by their parents – although, significantly, Plato had changed his own name (originally Aristocles) – and then proceed to common nouns, the choice of which is lost in prehistory. In both cases, Hermogenes takes the conventionalist

view and Cratylus the naturalist one. But Cratylus sounds absurd in the proper-name context when he holds that Hermogenes isn't really his interlocutor's name because he isn't actually 'born of Hermes' (which would imply that he is lucky and eloquent, when in fact he is neither). And Hermogenes sounds equally absurd in the common-noun context when he holds that it would make no difference if someone referred to a man as 'horse' and to a horse as 'man.'

Socrates raises the objection that, were this latter argument the case, there would be no way of distinguishing truth from falsehood, a position taken by relativists like Protagoras but anathema to Socrates (and Plato), for whom things have a natural reality of their own that does not depend on the perception of individuals. Socrates lays out the characteristically Platonic view that the things we perceive in the world around us are not really 'real,' in the sense that they are not permanent but in flux, are not the wholes we perceive but conglomerations of atoms, and are not perceived in the same way by all of us. The table at which I am now writing was not a table 30 years ago but has been made up from what was then parts of trees, which themselves did not exist 100 years ago, and in time the table will burn or decay, its substance taking on yet another form. The meaning of the word *table*, the knowledge I have of what a 'table' is, is not this transient thing. On the contrary, actual tables are made in accordance with the *function* of a table, which requires it to have a particular shape, that of a flat top surface supported by legs.

This is what Plato calls the *idea* of a table, what defines the *ideal form* of a table. That ideal form is the true, unchanging reality. Any individual table is merely an attempt to realize that ideal form in transient matter. The meaning of the word *table* is that idea, not any particular material instantiation of it. Moreover, ideas are all that we can 'know,' since knowledge must be of permanent, unchanging things, as opposed to the perception or opinion we can have of material things. We may disagree over whether a particular table is beautiful, but we cannot *know* that it is or is not beautiful in the same way that we can know what a table is and what beauty is – although, actually, such knowledge is not accessible to just anyone. The Ideal Forms inhabit a heaven into which only the philosopher, the wisest of men, can see (hence the political stance of Plato's *Republic*, where the ideal ruler is the philosopher-king).

From this point on, Socrates directs all his questions to Hermogenes, with Cratylus silent until quite late in the dialogue. In response to Socrates's point about truth, Hermogenes raises a powerful objection: If truth depends on some kind of natural relationship between word and thing, how is it possible for

different languages to exist? Socrates does not attempt to answer this directly (a sign of how seriously Plato took the question), but steers the dialogue off in the direction that the question demands. He asks Hermogenes about the purpose of words, and they conclude that words exist for two reasons: to discriminate among things, i.e., to pick out their *ousia*, the true essence that belongs to them alone; and to transmit that knowledge from the few who can perceive it directly to the many who cannot.

This leads Socrates to ask about the origins of the words we use, paving the way to the etymological inquiry that will form the great central bulk of the dialogue. Etymology, which in Greek means the 'study of truth,' was one of the 'sciences' based on language in which instruction was offered by Socrates's contemporaries the Sophists. The most successful and widely sought after of these was rhetoric, the art of using language in order to persuade, persuasion being the ultimate political commodity in the democracy of Athens. Just as they had no faith in democracy, Socrates and Plato disdained rhetoric as mere word-play, not at all concerned with real knowledge. Etymology, however, had more of an appeal. It not only claimed to be the study of truth, but did offer real insights into what words meant in an earlier time, closer to the moment of their creation. The great question, Socrates says, is whether whoever made the words we use – the semimythical *nomothetes*, which also means 'lawgiver' – really perceived the true essence of the thing he was naming, and if so, whether he succeeded in the word maker's craft of *mimesis*, 'imitation' of that essence in the sounds of language. This provides, in theory at least, the answer to Hermogenes's question about how it is possible for different languages to exist unless words are purely conventional: any number of 'correct' words are conceivable to designate a given meaning as long as they capture its essence and make it plain.

As Socrates proceeds through various classes of words and their etymologies, the thesis emerges that the creation of the Greek lexicon took place under the influence of Heraclitus's doctrine that 'everything flows.' Accounts are given whereby words of a positive moral sense are traced back to roots and sounds indicating motion, while those of a negative moral sense have roots expressing immobility. But all sorts of problems ensue. Socrates points out how easy it is, when no such account suggests itself, to have recourse to some *ad hoc* explanation, such as attributing a word to foreign origin. When it comes to *mimesis*, he notes that *r*, a trilled consonant in Greek, seems naturally to denote motion by the very way it is produced, and indeed appears in many

of the words with positive qualities that he traces back to the idea of motion – yet it does not appear in the word for ‘motion’ itself, *kinesis*. Socrates also cites cases of words known to have undergone sound change, which he puts down to people who “care nothing about the truth, only about how they shape their mouths.” This is a swipe not just at the *hoi polloi* in general but especially at rhetoricians and the poets who along with them are banned from Plato’s ideal Republic.

He finally turns back to Cratylus to discuss another fundamental problem with words being naturally connected to their meaning. Cratylus has insisted that the meaning of a ‘correct’ word must not only be embodied directly in the sounds of the word but also be indistinguishable from the idea of what it designates. Socrates leads him to admit that this is not in fact the nature of mimesis – if the image of something were identical to that something, it would not be an image at all but would actually have become the thing itself. Instead, the art of mimesis in language is to capture and reproduce some part of the essence of things; indeed, Socrates affirms, it is precisely in this way that a language *should* be constructed and used. But he admits, once constructed, the language passes into the hands of the many, who care only about the ‘vulgar’ function of communicating with their fellows, for which conventional words (not ‘correct’ ones in Cratylus’s sense) suffice. The investigation reaches an impasse, and Socrates concludes by opining that the Greek language really was created under the Heraclitean philosophy, but that the philosophy itself is misguided; that the lawgiver did not consistently embody its principles in sound; and that, even when he did so, words have subsequently undergone change at the hands of the mob, who care only about communication and sounding nice.

As a result, the dialogue ends up rejecting the possibility that the study of language opens a path to understanding the true nature of the universe. For each word, one would first have to decide whether it was created with an eye to truth, whether that truth was properly and consistently imitated in sounds, and whether the original form of the sounds has remained unaltered. But the first of these decisions already demands that we know what the truth is independently of the word in question – and if we can do that, there is no point inquiring further. This is clearly disappointing to Socrates, who believes that words should be naturally and deterministically bound to their meanings, but he is forced to admit that, at least as they are used for the vulgar purpose of communication, they are not in fact so bound. The way to knowledge of the truth is not therefore through the

study of etymology but through the kind of dialectic inquiry that the *Cratylus* itself embodies.

This is a blinding condemnation of the futility of any language-based ‘science.’ Yet so intense and sustained is Socrates’s attempt to disprove Hermogenes’s conventionalist stance through an examination of the hidden etymological structure of words that, although his attempt is ultimately given up as unsuccessful, the dialogue nevertheless has been read for centuries as a detailed study of Greek etymology. It was even interpreted as coming down on the side of Cratylus, holding that language really is grounded in nature, when in fact it makes clear that the positions taken by both Cratylus and Hermogenes are unsustainable in isolation from one another and from deeper inquiry into the nature of word making and mimesis.

The misinterpretations are less surprising when one considers how powerful and universal is the impulse to interpret language as having a meaning far deeper than what appears on the surface. This is nowhere more true than in European culture, formed by the doctrine of a god who is Himself *logos*, the Word, the language-based and language-like intelligence that created the world and continues to order it. From medieval through modern linguistic thought we find various attempts at containing arbitrariness, for instance, by locating it in nouns only, as opposed to verbs and other more ‘functional’ parts of speech; in the lexicon as a whole, as opposed to grammar; in conventional signs, which are imagined as having come about as an accretion onto more original and basic ‘natural’ signs; in a ‘universal’ grammar that underlies all languages, which differ from one another only in relatively superficial particulars.

Modern secular culture has not given up its faith in language but rather has problematized and debated its centrality in various guises that can be read as reworkings of the problem outlined and investigated in the *Cratylus*. Modern linguistics is predicated upon the belief in something deeper than the actual words we speak and hear, wherein their *real* structure lies, and perhaps also their real meaning. We find this impulse in Humboldt’s *energeia*, in Jakobson’s functionalism, in Chomsky’s universalism, and in various attempts to portray thought as conditioned by meaning, including the Sapir–Whorf hypothesis at the micro level of words, and the regimes of language described by present-day linguistic anthropologists and discourse analysts (in the wake of Nietzsche and Foucault) at the macro level of discourse. Even Saussure, with whom the modern doctrine of the arbitrariness of linguistic signs is most closely associated, taught that everything having to do with the systematic nature of languages – and languages

are entirely systematic by his account – needs to be approached from the point of view of limiting the arbitrary.

The legacy of the *Cratylus* has yet to play itself out. Once linguists manage to confront the powerful challenge it puts to the whole enterprise, we can perhaps begin to approach an understanding of language complex enough to take account of how Hermogenes, Cratylus, and Socrates are actually each right and wrong, and then to formulate, for the first time, new questions that actually pass beyond the concerns of the ancient dialogue rather than endlessly recycling them.

See also: Naturalism; Nominalism; Plato and His Predecessors; Realism and Antirealism.

Bibliography

- Baxter T M S (1992). *The Cratylus: Plato's critique of naming*. Leiden: E. J. Brill.
- Joseph J E (2000). *Limiting the arbitrary: linguistic naturalism and its opposites in Plato's Cratylus and modern theories of language*. Amsterdam & Philadelphia: John Benjamins.
- Kretzmann N (1967). 'History of semantics.' In Edwards P (ed.) *Encyclopedia of philosophy*, vol. 7. New York: Macmillan. 358–406.
- Sedley D (1998). 'The etymologies in Plato's *Cratylus*.' *Journal of Hellenic Studies* 118, 140–154.
- Sluiter I (1997). 'The Greek tradition.' In van Bakkum W, Houben J, Sluiter I & Versteegh K (eds.) *The emergence of semantics in four linguistic traditions: Hebrew, Sanskrit, Greek, Arabic*. Amsterdam & Philadelphia: John Benjamins. 147–224.

Plurality

P Lasersohn, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Plural expressions may be intuitively characterized as those involving reference to multiple objects. Semantic theories differ, however, in how this intuition is worked out formally.

The most popular approach is to treat plural expressions as referring to some sort of 'plural object,' or group. That is, alongside individual objects such as people, tables, chairs, etc., we assume there are groups of such objects, and that plural expressions refer to these groups in much the same way as singular expressions refer to individuals. Just as singular noun phrases denote and quantify over individuals, plural noun phrases denote and quantify over groups; just as singular predicates hold true or false of individuals, plural predicates hold true or false of groups.

An alternative, advanced by Schein (1993) (building on earlier logical work by George Boolos) is to regard a plural term as denoting each of several individuals directly, rather than denoting the group containing these individuals. This gives the effect of treating denotation for plural terms as a relation rather than a function; it is the denoting relation itself, rather than the denoted object, which is plural. The primary advantage of this technique is that it allows a reasonable treatment of noun phrases such as *the sets that do not contain themselves*, which potentially give rise to Russell's paradox in more conventional treatments.

Among analyses that do regard plural expressions as referring to groups, the main options are to identify groups with *sets*, or with *mereological sums*. The latter choice is favored especially by those who regard sets as abstract, mathematical objects existing outside of space and time. As Link (1998) puts it, "If my kids turn the living room into a mess I find it hard to believe that a set has been at work." However, not all authors share the intuition that sets of concrete objects are themselves abstract, and in any case the issue seems more philosophical than linguistic (but see the discussion of conjoined noun phrases below).

No matter which approach is adopted, a central problem in the semantics of plurality is determining the range and distribution of readings available to sentences containing plural expressions. Sentence (1a), for example, is intuitively interpreted as predicating the property of being numerous to our problems *collectively*; no individual problem can be numerous. Sentence (1b), in contrast, requires all (or nearly all) the individual children to be asleep; the predicate applies *distributively*. Sentence (1c) seems ambiguous between collective and distributive readings, meaning either that the T.A.s together earned exactly \$20 000, or that they each did:

- (1a) Our problems are numerous.
- (1b) The children are asleep.
- (1c) The T.A.s earned exactly \$20 000.

Such examples raise several issues: Are plural expressions authentically ambiguous between collective and distributive readings, or are both interpretations covered under a single, very general meaning?

If there is an authentic ambiguity, is it a simple two-way ambiguity between collective and distributive readings, or are there other possibilities? What is the locus of the ambiguity: in the noun phrase, the predicate, or both?

In favor of the view that there is an authentic ambiguity, consider a situation in which there are two T.A.s, and each of them earned exactly \$10 000. In this case, sentence (1c) is true. Sentence (2) is also true in this situation:

- (2) The T.A.s earned exactly \$10 000.

This suggests that there are two distinct figures, both of which are the exact amount the T.A.s earned: in one sense they earned \$20 000, and in another sense they earned \$10 000 – but this appeal to multiple senses amounts to a claim of ambiguity.

An ambiguity is also suggested by patterns of anaphora, as argued by Roberts (1991). Sentence (3a) allows the continuation in (3b) if (3a) is interpreted as predicating $\lambda x \exists y [\text{piano}(y) \ \& \ \text{lift}(x, y)]$ of the group of students as a whole, but not if this predicate is understood as applying to each student separately. (A third interpretation, in which there is some piano y such that the predicate *lifted* y applies to the individual students, also allows the continuation in (3b), but this does not affect the argument.)

- (3a) The students lifted a piano.
(3b) It was heavy.

If sentence (3a) is unambiguous, with no formal differentiation between the collective and distributive interpretations, it is hard to see how we could capture this difference in anaphoric potential.

If there is an ambiguity, one may ask whether the fully collective and fully distributive readings are the only ones available, or if instead there may be ‘intermediate’ readings. Intermediate readings appear to be called for in examples like (4):

- (4) The shoes cost \$75.

This sentence is most naturally interpreted as meaning that each pair of shoes costs \$75, not that each individual shoe costs that much, or that all the shoes together cost that much.

Gillon (1987) argues that sentences with plural subjects have as many readings as there are *minimal covers* of the set denoted by the subject noun phrase, in which a *cover* of a set A is a set C of nonempty subsets of A such that their union, $\cup C$, is equal to A , and a cover of A is *minimal* iff it has no subsets that are also covers of A ; this idea is developed further by Schwarzschild (1996) and others. Under this proposal, the pragmatic context makes a particular cover salient, and the predicate is required to hold of each

element of the cover. The fully distributive and fully collective readings reemerge as special cases.

However, cover-based analyses face a challenge in dealing with examples like (1c): Suppose John, Mary, and Bill are the T.A.s, and each of them earned \$10 000. In this case, the predicate *earned exactly* \$20 000 holds of each cell of the cover $\{\{\text{John, Mary}\}, \{\text{John, Bill}\}\}$, but sentence (1c) is not intuitively true in this situation.

Whether it appeals to covers or not, an ambiguity analysis must address the issue of where in the sentence the ambiguity is located. Early treatments often seemed to take for granted that it was the plural noun phrases themselves which are ambiguous, but many more recent treatments trace the ambiguity to the predicates with which the noun phrases combine (Scha, 1984; Dowty, 1986; Roberts, 1991; Lasersohn, 1995). A standard argument for this view comes from examples like (5):

- (5) The students met in the bar and had a beer.

The natural interpretation is that the students met collectively but had separate beers. This reading is easily obtained if we treat the subject noun phrase as unambiguously denoting the group of students as a whole, and treat the conjunct verb phrases as predicates of groups, with the distributive predicate *had a beer* holding of a group iff each of its individual members had a beer; the denotation of the whole, coordinate verb phrase may then be obtained by intersection. If we try to claim that the collective/distributive ambiguity is located in the subject noun phrase, however, it seems impossible to give a consistent answer as to which reading it takes in this example.

A suitable ambiguity in the predicate may be obtained by positing an implicit operator on the predicate. Link (1998) and Roberts (1987) suggest an optional operator on the verb phrase, notated “ D ”, and interpreted as $\lambda P \lambda x \exists y [y \Pi x \rightarrow P(y)]$, where Π is the relation an individual stands in to the larger groups of which it forms a part. This gives a simple two-way ambiguity, depending on whether the operator is present or not; intermediate readings may be obtained using a more complex operator which quantifies over elements of a cover (Schwarzschild, 1996). Either operator is easily generalized across types to give distributive readings for non-subject argument places (Lasersohn, 1998).

It should be noted that the lexical semantics of some predicates prevent them from participating in the collective/distributive ambiguity: a verb like *sleep*, for example, cannot apply to a group without also applying to the individual members of the group. Adding a distributivity operator in this case is redundant, and does not result in a difference in meaning.

Conversely, certain predicates apply only to groups: *gather*, for example. Such predicates do show collective/distributive ambiguities however, when applied to arguments denoting groups of groups:

- (6) The tribes gathered.

Example (6) may be used to mean that each tribe gathered separately, or that all the tribes gathered together.

A good deal of work has been devoted to the relation between *conjunction* and plurality (Link, 1998; Hoeksema, 1983; Landman, 1989a; Lasersohn, 1995; Schwarzschild, 1996; Winter, 2001) because conjoined noun phrases sometimes admit collective readings, as in (7):

- (7) John and Mary lifted the piano.

Such examples suggest that coordinate noun phrases may denote groups in much the same fashion as plural noun phrases. Such sentences also admit a distributive reading, which may be obtained either through the use of a distributivity operator as with other plural noun phrases, or by reducing the coordination to propositional conjunction, via a generalized conjunction operator or conjunction reduction transformation.

Conjoined noun phrases have sometimes been used to argue that semantic theory must allow reference to *higher-order groups* (Hoeksema, 1983; Landman, 1989a). This is easily accomplished in set theory, since sets may contain other sets as members: $\{a, \{b, c\}\} \neq \{\{a, b\}, c\} \neq \{a, b, c\}$, but is less straightforward if groups are modeled as mereological sums.

- (8a) Blücher and Wellington and Napoleon fought
against each other at Waterloo.
(8b) The cards below seven and the cards from seven
up are separated.

Example (8a), from Hoeksema (1983), may be parsed in either of two ways: $[[\text{Blücher and Wellington}] \text{ and } [\text{Napoleon}]]$ or $[[\text{Blücher and } [\text{Wellington and Napoleon}]]]$. It is intuitively true relative to the first parse but false relative to the second, suggesting that the two parses correspond to denotations such as $\{\{b, w\}, n\}$ and $\{b, \{w, n\}\}$. Likewise example (8b), from Landman (1989a), is not equivalent to *The cards below 10 and the cards from 10 up are separated*. But if reference to higher-order groups is disallowed, the subject noun phrases of these two sentences would seem to refer to the same group, namely the group containing all the individual cards as members. The opposing view that noun phrases need never refer to higher-order groups has been defended in detail by Schwarzschild (1996), who points out that the pragmatic context may make

salient a particular division of the denotation into subgroups even if that denotation is first-order; correct truth conditions may be obtained if the semantics is sensitive to this pragmatically supplied division.

A number of additional issues arise in the semantics of plurality, which can only briefly be mentioned here: Certain *adverbs*, such as *together* or *separately*, seem to force a collective or distributive reading, but the exact mechanism by which they do this is a matter of some dispute (Lasersohn, 1995; Schwarzschild, 1994; Moltmann, 1997). Plural expressions share a number of characteristics with *mass terms*, suggesting a unified analysis (Link, 1998). Plural noun phrases affect the *aspectual class* of predicates with which they combine, suggesting a parallel in the domain of *events* to the structure of groups among individuals (Krifka, 1989) – a parallel also suggested by the phenomenon of *verbal plurality* or ‘pluractionality’ (Lasersohn, 1995). Finally, the interpretation of *bare plurals*, or plural noun phrases with no overt determiner, and particularly the alternation between the *existential*, *generic*, and *kind-level* readings, illustrated in (9), has attracted enormous attention; but much of this work belongs more properly to the study of genericity and indefiniteness than to the semantics of plurality *per se* (Carlson, 1980; Carlson and Pelletier, 1995).

- (9a) Raccoons are stealing my corn.
(9b) Raccoons are sneaky.
(9c) Raccoons are widespread.

See also: Generic Reference; Mass Nouns, Count Nouns, and Non-count Nouns: Philosophical Aspects.

Bibliography

- Carlson G (1980). *Reference to kinds in English*. New York: Garland Press.
Carlson G & Pelletier F J (1995). *The generic book*. Chicago: University of Chicago Press.
Dowty D (1986). ‘Collective predicates, distributive predicates, and *all*.’ In Marshall F (ed.) *ESCOL '86: Proceedings of the Third Eastern States Conference on Linguistics*. Columbus: Ohio State University. 97–115.
Gillon B (1987). ‘The readings of plural noun phrases in English.’ *Linguistics and Philosophy* 10, 199–219.
Hamm F & Hinrichs E (1998). *Plurality and quantification*. Dordrecht: Kluwer Academic Publishers.
Hoeksema J (1983). ‘Plurality and conjunction.’ In ter Meulen A (ed.) *Studies in model theoretic semantics*. Dordrecht: Foris Publications. 63–83.
Krifka M (1989). ‘Nominal reference, temporal constitution, and quantification in event semantics.’ In Bartsch R, van Benthem J & van Emde Boas P (eds.)

- Semantics and contextual expression*. Dordrecht: Foris Publications. 75–115.
- Landman F (1989a). ‘Groups, I.’ *Linguistics and Philosophy* 12, 559–605.
- Landman F (1989b). ‘Groups, II.’ *Linguistics and Philosophy* 12, 723–744.
- Landman F (2000). *Events and plurality: the Jerusalem lectures*. Dordrecht: Kluwer Academic Publishers.
- Lasnik P (1995). *Plurality, conjunction and events*. Dordrecht: Kluwer Academic Publishers.
- Lasnik P (1998). ‘Generalized distributivity operators.’ *Linguistics and Philosophy* 21, 83–92.
- Link G (1998). *Algebraic semantics in language and philosophy*. Stanford, California: CSLI Publications.
- Moltmann F (1997). *Parts and wholes in semantics*. Oxford: Oxford University Press.
- Roberts C (1991). *Modal subordination, anaphora and distributivity*. New York: Garland Press.
- Scha R (1984). ‘Distributive, collective and cumulative quantification.’ In Groenendijk J, Janssen T & Stokhof M (eds.) *Truth, interpretation and information*. Dordrecht: Foris Publications. 131–158.
- Schein B (1993). *Plurals and Events*. Cambridge, Mass.: MIT Press.
- Schwarzschild R (1994). ‘Plurals, presuppositions, and the sources of distributivity.’ *Natural Language Semantics* 2, 201–248.
- Schwarzschild R (1996). *Pluralities*. Dordrecht: Kluwer Academic Publishers.
- Winter Y (2001). *Flexibility principles in boolean semantics*. Cambridge, Mass.: MIT Press.

Polysemy and Homonymy

A Koskela and M L Murphy, University of Sussex, Brighton, UK

© 2006 Elsevier Ltd. All rights reserved.

Polysemy and homonymy both involve the association of a particular linguistic form with multiple meanings, thus giving rise to **lexical ambiguity**.

Polysemy is rooted in a variety of semantic-pragmatic processes or relations through which meanings of words extend or shift so that a single lexical item (a **polyseme**) has several distinct senses. For example, *language* is **polysemous** in that it can be used to refer to the human linguistic capacity (*Language evolved gradually*) or to a particular grammar and lexis (*Learn a new language!*). The most clear-cut cases of polysemy (versus homonymy) involve **systematic** (or **regular**) **polysemy** (Apresjan, 1974), in which the relation between the senses is predictable in that any word of a particular semantic class potentially has the same variety of meanings. For example, words for openable coverings of apertures in built structures (*She rested against the door/gate/window*) are also used to refer to the aperture itself (*Go through the door/gate/window*). In **non-systematic** polysemy, the word’s two senses are semantically related, but are not part of a larger pattern, as for *arm* of government versus human *arm*. Within the literature, theoretical considerations often lead authors to use *polysemy* to refer only to either systematic or non-systematic polysemy, and so the term must be approached with caution.

Homonyms, in contrast, are distinct lexemes that happen to share the same form. They arise either accidentally through phonological change or lexical

borrowing, or through some semantic or morphological drift such that a previously polysemous form is no longer perceived as being ‘the same word’ in all its senses. *Tattoo*₁ ‘an ink drawing in the skin’ and *tattoo*₂ ‘a military drum signal calling soldiers back to their quarters’ provide a clear example of homonymy, in that the two words derive respectively from Polynesian *ta-tau* and Dutch *taptoe* ‘turn off the tap’ (which was also used idiomatically to mean ‘to stop’). The formal identity of homonyms can involve both the phonological and the written form, as for *tattoo*. **Homophones** need only be pronounced the same, as in *tail* and *tale*, and **homographs** share written form, but not necessarily phonological form, as in *wind* /wind/ and *wind* /waɪnd/. Homographs that are not homophones are also called **heteronyms**.

Polysemy and homonymy are generally differentiated in terms of whether the meanings are related or not: while polysemous lexical items involve a number of related meanings, homonyms, as accidentally similar words, do not have any semantic relation to each other. However, as discussed below, a clear distinction between polysemous and homonymous items remains difficult to draw. Nevertheless, the need to distinguish between them remains. For lexicographers, the distinction generally determines how many entries a dictionary has – homonyms are treated as multiple entries, but all of a polyseme’s senses are treated in one entry. For semanticists wishing to discover constraints on or processes resulting in polysemy, weeding out the homonymous cases is necessary. For those wishing to model the mental representation of lexical knowledge, the issue of semantic (un)relatedness is similarly important.

Another definitional issue concerns the distinction between polysemy and **vagueness**, in which a lexical item has only one general sense. (Some authors refer to this as **monosemy**.) This raises the question of how different two usages of the same form need to be to count as distinct polysemous senses rather than different instantiations of a single underlying sense. For example, the word *cousin* can refer to either a male or a female, but most speakers (and linguists) would not view *cousin* as having distinct ‘male cousin’ and ‘female cousin’ senses. Instead, we regard it as vague with respect to gender. A number of different methodologies and criteria have been used to draw the line between polysemy and vagueness, but these are not uncontroversial. The distinction is also influenced by theoretical assumptions about the nature of lexical meaning; thus, demarcation of senses is one of the more controversial issues in lexical semantics.

Evidence Used in Differentiating Homonyms and Polysemes

The kinds of evidence used for differentiating polysemous and homonymous items can be roughly divided into two types: evidence regarding the relatedness of the meanings involved and evidence regarding any formal (morphosyntactic or phonological) differences in the linguistic form that correspond to the distinct senses.

The key principle for distinguishing polysemes and homonyms involves the relatedness of the meanings associated with the linguistic form: unrelated meanings indicate homonymy whereas related meanings imply polysemy. Relatedness can either be determined diachronically by establishing if the senses have a common historical origin, or synchronically – and neither method is unproblematic. The simplest criterion for synchronic relatedness is whether or not the senses participate in the same semantic field. On this criterion, metonymically related senses such as *bench* ‘place where the judge sits’ and ‘the judge’ are polysemous, but metaphorically related senses such as *foot* ‘body part’ and ‘bed part’ are not. Nevertheless, the two senses of *foot* here are usually considered to represent polysemy, as the more usual methodology for determining synchronic relatedness involves individuals’ intuitions. Intuitive judgements of semantic relations are, however, always subjective, and it may be questioned whether people’s metalinguistic reasoning about meaning relations can reflect how lexical meaning is mentally represented. Historical evidence has in its favor the fact that etymological relations are often less equivocal; but many approaches consider such evidence irrelevant to

questions of mental representation, as etymological information is not part of most speakers’ competence. Furthermore, diachronic and synchronic evidence can be contradictory. In some cases, despite the fact that the different senses are etymologically related, native speakers today perceive no semantic relation. Such is the case with the classic homonym *bank*; the two senses ‘financial institution’ and ‘a raised ridge of ground’ can be traced back to the same proto-Germanic origin. There are also instances where at least some speakers perceive a synchronic relation in words with no shared history. For example, *ear* ‘hearing organ’ and *ear (of corn)* are perceived by some speakers to be metaphorically related, analogous to other body-part metaphors, but in fact they have separate sources.

Formal criteria are also used to distinguish homonymy and polysemy. One criterion, adopted widely in lexicography, is that polysemous senses should belong to the same grammatical category. Thus, noun and verb senses of words such as *waltz*, derived through zero derivation (conversion), might be considered distinct lexical items, and therefore homonyms. Similarly, the potential for different plural forms for the meanings of *mouse*, ‘small rodent’ and ‘a computer input device’ (*mice* and *mouses*, respectively), could indicate that the distinct senses reflect distinct homonymous lexemes. However, the formal criterion often conflicts with the semantic criteria. One can easily see the semantic connection between *waltz* (n.) and *waltz* (v.) as part of a pattern of regular polysemy (where any noun denoting a type of dance can also be a verb meaning ‘to perform that dance’), and most people appreciate the metaphorical connection between the two types of *mouse*. But for some approaches to the lexicon, any morphological differences associated with different senses of the form force its treatment as a homonym, resulting in etymologically related homonyms. The contradictions of the formal, diachronic, and synchronic semantic criteria have led some to abandon strict distinctions between homonyms, polysemes, and vague lexemes. Tuggy (1993) proposes a continuum between these categories, which relies on variable strengths of association among meanings and forms.

Theoretical Approaches to Polysemy and Homonymy

While we have so far attempted a theory-neutral definition and description of polysemy and homonymy, variability in the interpretation of these terms is often at the heart of the contrasts among theoretical approaches. The approach is often determined by the

range of phenomena considered (and *vice versa*) – for instance some models define *polysemy* as equivalent to *systematic polysemy* and treat irregular cases as tantamount to homonymy.

In early generative theory, attempts were made to treat polysemy as the product of synchronic derivational processes. The lexical representation of a word would include a meaning, and further meanings could be derived via lexical rules that operate on some semantic class of words (e.g., McCawley, 1968; Leech, 1974). Leech (1974) notes that such rules are limited in their productivity, and can be seen as motivating but not predicting new senses of words. The assumption of a basic meaning from which additional meanings are derived continues as a theme in some **pragmatic** approaches to polysemy. For instance, Nunberg (1978) posits that words have lexical meanings but that they can be used to refer in various ways based on a number of conventional referring functions that allow language users to effect different sense extensions. Thus, a referring function that relates producers with their products predicts the interpretation of *Chomsky* (primary reference is to a person) as ‘the works of Chomsky’ in *Chomsky is hard to read*. Other theorists have taken the position that word meaning is radically underspecified in the lexicon (e.g., Bierwisch, 1983) or extremely general in sense (Ruhl, 1989) and that semantic or pragmatic factors allow for more specific interpretations in context. Such an approach is Blutner’s (1998) *Lexical Pragmatics*, where lexical meaning is highly underspecified. Meanings are enriched by a pragmatic mechanism that specifies which particular concepts the word refers to in a particular use based on contextual factors. Such approaches erase the distinction between polysemy and vagueness.

The **generative lexicon** (Pustejovsky, 1995) is an approach to lexical semantics that emerges from computational linguistic work. Sense disambiguation presents a key challenge for natural language processing and drives much current polysemy research. In this theory, systematic polysemy is generated by lexical rules of composition that operate on semantic components specified in the representation of lexical items (and so it could be seen as a development from the early generative approaches). The main concern is to account for regular types of meaning alternation (such as the aperture-covering alternation of *door* and adjectival meaning variation, as in *fast car*, *fast typist*, *fast decision*, *fast road*, etc.) in a manner that avoids the problems of simple enumeration of word senses in the lexicon and explicates the systematic nature of meaning variation in relation to the syntagmatic linguistic environment. The meaning variation of *fast*,

for example, is accounted for by treating the adjective as an event predicate which modifies an event that is specified as the head noun’s function as part of its compositional lexical representation.

Another significant strand of polysemy research proceeds within **Cognitive Linguistic** approaches. The general aim is to study the kinds of conceptual processes that motivate the multiple meanings of linguistic forms and how these meanings may be grounded in human experience. As it is argued that lexical categories exhibit the same kind of prototype structure as other conceptual categories, the relations of polysemous senses are usually modeled in terms of radial, family resemblance categories. In these polysemy networks, the senses are typically either directly or indirectly related to a prototypical sense through such meaning extensions processes as conceptual metaphor and metonymy and image-schema transformations (Lakoff, 1987). As Cognitive Linguists also assume that all of linguistic structure, including grammar, is meaningful, some work within this approach has extended the applicability of the notions of polysemy and homonymy beyond lexical semantics to grammatical categories and constructions (Goldberg, 1995). While most other approaches dismiss homonymy as accidental and uninteresting, the relevance of diachronic processes to meaning representation in Cognitive Linguistics means that making hard and fast distinctions between homonymy, polysemy, and vagueness is not necessary (Tuggy, 1993; Geeraerts, 1993).

Polysemy and homonymy thus represent two central notions in lexical semantics, inspiring active research interest.

See also: Lexical Semantics: Overview.

Bibliography

- Apresjan J (1974). ‘Regular polysemy.’ *Linguistics* 142, 5–33.
- Bierwisch M (1983). ‘Semantische und konzeptuelle Repräsentation lexikalischer Einheiten.’ In Růžicka R & Motsch W (eds.) *Untersuchungen zur Semantik*. Berlin: Akademie-Verlag.
- Blutner R (1998). ‘Lexical pragmatics.’ *Journal of Semantics* 15, 115–162.
- Geeraerts D (1993). ‘Vagueness’s puzzles, polysemy’s vagaries.’ *Cognitive Linguistics* 4, 223–272.
- Goldberg A E (1995). *Constructions*. Chicago: University of Chicago Press.
- Lakoff G (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
- Leech G (1974). *Semantics*. Harmondsworth: Penguin.

- McCawley J D (1968). 'The role of semantics in a grammar.' In Bach E & Harms R T (eds.) *Universals in linguistic theory*. New York: Holt, Rinehart & Winston. 124–169.
- Nunberg G (1978). *The pragmatics of reference*. Bloomington, IN: Indiana University Linguistics Club.
- Pethő G (2001). 'What is polysemy?' In Németh E & Bibok K (eds.) *Pragmatics and the flexibility of word meaning*. Amsterdam: Elsevier. 175–224.

- Pustejovsky J (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Ravin Y & Leacock C (eds.) (2000). *Polysemy*. Oxford: OUP.
- Ruhl C (1989). *On monosemy*. Albany: SUNY Press.
- Tuggy D (1993). 'Ambiguity, polysemy and vagueness.' *Cognitive Linguistics* 4, 273–290.

Possible Worlds: Philosophical Theories

D Gregory, University of Sheffield, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

Anyone who reads contemporary philosophy will soon encounter talk of possible worlds, and anybody who delves a little deeper will quickly discover that there are various conflicting philosophical accounts of their nature. We consider those competing accounts, but it is helpful to begin with a discussion of the ends that possible worlds are commonly made to serve.

Possible worlds made their debut on the current philosophical scene through Kripke's work on the metamathematics of modal logics, but their current ubiquity probably owes more to their relationships to less esoteric matters. We commonly express claims about what is possible by using statements that appear to assert the existence of 'possibilities,' for example, or of 'ways things might have been.' So, for instance, somebody might assert, quite unexceptionably, that 'there's a possibility that the world will end in the next 10 years.'

That habit has striking affinities with one of the characteristic principles concerning worlds, that what is possible at a given world w is what obtains at some world that is possible relative to w . Another central principle invoking worlds relates to necessity: what is necessary at a given world w is what holds at every world that is possible relative to w . That last principle in fact follows immediately from the previous one (and vice versa) on the assumption that worlds are *complete*, in the sense that each proposition is either true at a given world or false at it.

The above theses may be interpreted in numerous ways, depending upon how their talk of 'possibility,' 'necessity,' and 'possible worlds' is read. For instance, the first principle can be construed as concerning the *physical* possibilities at a given world – whatever is compatible with the fundamental physical nature of that world – so long as we restrict the 'worlds' there considered to those that are *physically* possible relative to the relevant possible world. And if we impose

that restriction upon the possible worlds cited in the second principle, it can be treated as applying to physical necessities.

Another central assumption involving possible worlds is that one among them is the *actual* world, at which precisely the actual truths obtain. That assumption combines with the various readings of the earlier theses relating to possibility and necessity to yield truth conditions for a wide variety of modal statements. So, for instance, it implies that it is actually physically necessary that P just in case P is physically necessary at the actual world. But the central hypothesis relating possible worlds to physical necessities implies that P is physically necessary at the actual world just in case P holds at every world that is physically possible relative to the actual world.

The various theses just considered are perhaps the least contentious principles featuring possible worlds (another, more contentious but widely accepted use for worlds is in providing truth conditions for counterfactual conditionals, like 'if tigers had 10 legs they would be cumbersome'). Elsewhere, contention is the order of the day. Thus some philosophers, like David Lewis, have claimed that possible worlds can be used to provide thoroughly nonmodal analyses of modal claims. But others, like Alvin Plantinga, have disagreed. And some philosophers – Lewis again, for instance – hope to reduce propositions to classes of worlds, whereas others prefer to follow Robert Adams and Arthur Prior in identifying possible worlds with special sorts of propositions, or with set-theoretical constructions based upon propositions.

There is disagreement, then, over what it is reasonable to expect from possible worlds – that is, over what we can sensibly hope to use them *for*. Those differences are echoed in the varying theories of what possible worlds are. For it is commonly held that the concept of a possible world is a *functional* one: possible worlds are those things that are fit to play certain roles in philosophical theorizing about modality and related matters. But if one philosopher thinks a certain range of roles should be filled by possible

worlds, while another believes that worlds should serve a somewhat different range of roles, their accounts of which things are possible worlds may consequently diverge simply because the two philosophers have focused on groups of jobs that call for different occupants.

At one end of the spectrum, Lewis attempted to make possible worlds perform an extraordinarily ambitious range of tasks (for a comprehensive outline of Lewis's views on possible worlds, see Lewis, 1986). Lewis claimed that a possible world is a group of things that are all spatiotemporally related to one another and where each thing that is spatiotemporally related to something in the group is also among its occupants. So, for instance, the actual world contains precisely those things that stand in a spatiotemporal relation to you or me.

Lewis used his nonmodal account of the nature of possible worlds to provide wholly nonmodal analyses of modal locutions. He also followed the common practice of using talk of possible worlds in formal semantical treatments of fragments of natural language. And he identified numerous types of entities that philosophers and nonphilosophers have posited, such as properties and propositions, with set-theoretical constructions founded on possible worlds and their inhabitants. Another important aspect of Lewis's position is his denial that distinct possible worlds ever share any inhabitants. This led him to develop *counterpart theory*, according to which a statement regarding a certain individual belonging to a given possible world *w* – 'Kant had a beard,' for instance – is true in another possible world *y* just in case *y* contains a bearded entity that is sufficiently similar to Kant.

Although Lewis argued with great virtuosity that his putative possible worlds would perform the various tasks that he considered and that we ought therefore to believe that his worlds exist, some of his theory's obvious consequences are so implausible that few people have been willing to accept his conclusions. For instance, there might have been talking donkeys. So, Lewis claimed, there is a group of spatiotemporally interrelated items that includes a talking donkey. Hence, a talking donkey exists even if none *actually* exists – that is, a talking donkey exists even if we do not stand in any spatiotemporal relations to such a beast.

Lewis placed great emphasis upon his theory's provision of nonmodal analyses of modal locutions and argued that none of his view's major competitors could also provide them. But some of his opponents would distance themselves from Lewis's reductionist aims. Indeed, a significant lacuna in Lewis's case for his position is that he never provided a compelling account of why we should want nonmodal analyses

of modal locutions. For unless one is already persuaded of the desirability of such analyses, one of the chief supposed virtues that Lewis claimed for his stance seems instead to be a mere curiosity.

Following van Inwagen (1986), Lewis's theory may be described as *concretist*, because his possible worlds and their inhabitants are concrete rather than abstract. Lewis's view thus contrasted with the preponderance of accounts of possible worlds, on which possible worlds are identified with paradigmatically abstract items of some kind or another. For example, Adams (1974) identified possible worlds with special sets of propositions where propositions were assumed to be a variety of abstract object; Plantinga (1974) identified possible worlds with a certain type of states of affairs, another putative type of abstract item; and Stalnaker (1976) identified worlds with a particular variety of ways things might have been, which he took to be a sort of property.

Theorists who identify possible worlds with abstract entities – *abstractionists*, to use some more of van Inwagen's terminology – need not endorse highly revisionary views concerning what concrete objects exist. That fact makes their general approach more immediately appealing than Lewis's concretism. Of course, if abstractionists are to respect our ordinary modal opinions, they must still posit very many abstract objects; each total possible state of the world should correspond to a possible world. But we appear to accept that the abstract domain is immensely populous (we seem to believe in infinite collections of numbers, for instance), and that fact may make us jib less, perhaps irrationally, at the hefty ontologies that abstractionists require as compared to equally large concretist ontologies.

Abstractionists nonetheless have some work to do if they are to make the ontological foundations of their theories credible. Russell's discovery that the guiding principles of naive set theory are inconsistent showed that one cannot be assured that there really are abstract items answering to every intuitively appealing map of a portion of the abstract realm. Abstractionists therefore need to persuade us that the abstract things with which they identify possible worlds exist. To take a specific case, why should we believe in the abstract states of affairs that Plantinga equates with possible worlds? At the least, Plantinga should present a strong case that the conception of states of affairs that underlies his approach is consistent. Similar remarks apply to the other abstractionists mentioned earlier, Adams and Stalnaker.

As stated earlier in this article, Lewis argued that abstractionist accounts of possible worlds cannot provide nonmodal analyses of modal statements. So what can abstractionists do with their putative

possible worlds? They can use them to supply truth conditions for many modal claims, for one thing, although those truth conditions may not be stated nonmodally. (On Adams's theory, for instance, the supplied truth conditions will speak of 'consistent' sets of propositions). And they can provide interpretations of those philosophical discussions, which, rather than address the question of what possible worlds are, instead take possible worlds for granted and proceed to frame modal arguments and theses by speaking of them. While those tasks may seem trifling when compared to the more spectacular reductionist uses for possible worlds proposed by Lewis, they should not be dismissed – as noted at the outset, the spread of possible worlds through recent philosophy is, after all, owed to their use in precisely those ways and not to a widespread conviction that the modal ultimately boils down to the nonmodal.

An apparently simple method, described clearly by Lewis at various points in his writings, has very often guided philosophical investigations into the nature of possible worlds: one lines up the various contending accounts; one then compares their costs and benefits; and one opts for the position that comes out best overall. But although that methodology looks straightforward, its correct application requires a prior determination of philosophical virtues and vices, and the latter task is not trivial. So, for instance, how are to decide whether Lewis's nonmodal theory of possible worlds should be preferred on that account to the modal theories frequently offered by abstractionists? And how are we to adjudicate the sometimes competing demands of commonsense modal opinion and theoretical elegance?

Those and similar questions have perhaps been a little neglected by philosophers of modality, and further investigation into their answers might breathe life into a debate that has lately looked somewhat stalled. Accounts of possible worlds proliferate, but attempts to figure out what precisely we should

demand from such theories are surprisingly scarce. This would be understandable if the discussion were to manifest a high degree of well-grounded consensus on the underlying desiderata, but that condition is evidently unsatisfied: what looks like a philosophical imperative to one philosopher – the need to avoid modal primitives of one kind or another, say, or to ensure that first-order modal logic has certain inferential features – can often appear a mere fetish to another. Of course, it may be that the current debate merely reflects the fact that we cannot realistically aim for wholly satisfying answers to the sort of questions just identified; but that prognosis is a dismaying one that we should not lightly accept.

See also: Counterfactuals; Formal Semantics; Modal Logic; Montague Semantics.

Bibliography

- Adams R M (1974). 'Theories of actuality.' *Nous* 8, 211–231. [Reprinted in Loux (1979).]
 Divers J (2002). *Possible worlds*. London: Routledge.
 Forbes G (1985). *The metaphysics of modality*. Oxford: Clarendon Press.
 Hughes G E & Cresswell M (1968). *A new introduction to modal logic*. London: Routledge.
 Kripke S (1980). *Naming and necessity*. Oxford: Blackwell.
 Lewis D (1986). *On the plurality of worlds*. Oxford: Blackwell.
 Loux M J (ed.) (1979). *The possible and the actual*. Ithaca: Cornell University Press.
 Plantinga A (1974). *The nature of necessity*. Oxford: Clarendon.
 Prior A N & Fine K (1977). *Worlds, times and selves*. London: Duckworth.
 Rosen G (1990). 'Modal fictionalism.' *Mind* 99, 327–354.
 Stalnaker R (1976). 'Possible worlds.' *Nous* 10, 65–75. [Reprinted in Loux (1979).]
 van Inwagen P (1986). 'Two concepts of possible worlds.' *Midwest Studies in Philosophy* 11, 185–213.

Pragmatic Determinants of What Is Said

E Borg, University of Reading, Reading, UK

© 2006 Elsevier Ltd. All rights reserved.

Imagine that Jack, pointing at a book on the table, utters the sentence "That is better than Jill's book." What, we might ask, is said by his utterance? Well, since the sentence contains a context-sensitive expression – the demonstrative 'that' – we know that

determining what is said will require an appeal to the context of utterance to provide a referent. Thus, this is one kind of pragmatic determinant of what is said by Jack's utterance. The sentence is also context sensitive in another, slightly less obvious, way: it is in the present tense, so an appeal to the context of utterance will also be needed to fix a time for Jack's claim. Finally, many theorists recently have argued that the sentence is also context sensitive in a third, much less

obvious, way: for to grasp what is said by Jack it seems that we also need to know what kind of relationship he envisages between Jill and her book. Does Jack intend to say that the book he is demonstrating is better than the book Jill wrote, better than the book Jill owns, the book Jill is reading, etc? And what does he mean by 'better' here – better written, better researched, better for standing on to reach a high shelf? Surely Jack says more than merely that the book he is demonstrating is better in some respect or other than a book bearing some relationship or other to Jill, but, if so, then any such additional information can only come from consideration of the context of utterance.

Furthermore – and this is where the current case differs from that of demonstratives and tense – it's not obvious that this last type of pragmatic determinant of what is said is traceable to any overt (i.e., syntactically represented) context-sensitive element. Jack's utterance does contain the syntactically marked possessive, but 's certainly doesn't appear on the list of usual suspects for context-sensitive expressions. And in many cases even this degree of syntactic representation is missing. Consider an utterance of "It's raining," where what is said is that it is raining at some location, I, or "You won't die," saying that you won't die from that cut, or "Nietzsche is nicer," meaning that Nietzsche is nicer than Heidegger. Or again, consider the speaker who produces a nonsentential utterance, e.g., pointing at a child and saying "George's brother," thereby conveying that he is George's brother. In all these cases we have additional information contributed by the context of utterance yet which seems unmarked at the syntactic level.

These types of pragmatic determinant of what is said (i.e., contextually supplied information which is semantically relevant but syntactically unmarked) have come to be known as 'unarticulated constituents' (UCs). Such elements are extremely interesting to philosophers of language since they seem to show that there are problems with a quite standard conception of formal semantics, according to which the route to meaning runs along exclusively syntactic trails.

Now, as noted above, even in formal theories of meaning, like truth-conditional accounts (*see Truth Conditional Semantics and Meaning*), some appeal to a context of utterance must be made in order to cope with overtly context-sensitive expressions. However, the thought has been that this need not contravene the essentially formal nature of our theory, since the contextual instructions can be syntactically triggered in this case (crudely, the word 'that' tells us to find a demonstrated object from the context of utterance). Much more problematic, then, are pragmatic determinants of what is said which apparently lack any

syntactic basis, for they show that context can come to figure at the semantic level regardless of whether it is syntactically called for (thus there will be no purely syntactic route to meaning).

In response to the challenge posed by UCs, advocates of formal semantics have sought to reject the idea that there are any such things as syntactically unrepresented but semantically relevant elements, claiming either that

1. every contextual element of what is said really is syntactically represented, even though this fact may not be obvious from the surface syntax of a sentence (i.e., there is more to our syntax than initially supposed), or that
2. any contextual contributions which are not syntactically marked are not semantically relevant (i.e., there is less to our semantics than initially supposed).

The first of these moves is pursued by a number of theorists who argue that, if one pays proper attention to all the information given at the syntactic level, one will find that many cases of supposedly syntax-independent semantic contribution are, in fact, syntactically required (see Stanley, 2000; Stanley and Szabo, 2000; Taylor, 2001; Recanati, 2002). Now, that there is some discrepancy between surface form and underlying syntax is a well-rehearsed point. For instance, Russell's theory of descriptions invokes a clear distinction between surface form and logical form. Furthermore cases of so-called syntactic ellipsis (like the contraction in "Jack likes dogs and so does Jill," where the second sentence is held to have the underlying syntactic form "Jill likes dogs" despite its reduced verbal form) show that surface constituents may be poor indicators of syntactic elements. If this is correct, then the principle that there is more to our syntax than is apparent at first glance is independently well motivated.

Given this, there certainly seem to be cases in which (1) is an appealing response to putative examples of UCs. To give an example: someone might claim that, because an utterance of "Jack plays" can be used in one context to say that Jack plays the trombone and in another to say that Jack plays football, there must be a syntactically unmarked, contextual contribution to the semantic content of an utterance of this kind. However, closer inspection of the syntax shows this conclusion is dubious: 'plays' is a transitive verb requiring a subject and an object and, although the argument place for an object can be left unfilled at the surface level without this making the sentence ill formed, it might be argued that the additional argument place is marked in the underlying syntax of the verb. Since 'plays' is a transitive verb, the competent interlocutor will, on hearing "Jack plays," expect to

look to a context of utterance to discover what Jack plays, but this is precisely because she is sensitive to the syntactic structure of English in this case. Thus sometimes an appeal to an enriched syntax seems the best way to cope with putative UCs. However, there are also cases where (1) seems less compelling. For instance, in certain contexts, it seems that Jack can use an utterance of “The apple is red” to convey the proposition that the apple is red to degree n on its skin. Yet it is far from clear that the correct syntactic structure for color terms includes argument places for the shade, or precise manner of instantiation, of the color. Furthermore, the idea that syntactic structure can outstrip surface form at all (or at least in the ways required by [1]) has been rejected by some.

So, the opponent of UCs might seek to supplement (1) with (2), allowing that some contextually supplied information is not marked at the syntactic level but denying that such information is semantically relevant. The advocate of this kind of response wants to accept that, at an intuitive level, the speaker who utters “The apple is red” (or “You won’t die,” etc.) clearly does convey the contextually enriched proposition (i.e., that the apple is red on the outside, or that you won’t die from that cut) but that this is an instance of speaker meaning rather than semantic content (*see Semantics–Pragmatics Boundary*). Now, whether this move is warranted in all cases is again something of a moot point. One objection might be that it is a mistake to seek to separate speaker intuitions about what is said and claims of semantic content to the radical degree predicted by this kind of move. So, for instance, it seems that the message recovered from an utterance of “That banana is green” will always be that that banana is green in some salient respect, but if this is the kind of proposition competent speakers can and do recover in communicative exchanges, shouldn’t this be what semantics seeks to capture? What role could there be for a more minimal kind of semantic content (e.g., which treats semantic content as exhausted by the proposition that that banana is green *simpliciter*)? (See Borg, 2004; Cappelen and Lepore, 2005, for attempts to answer this question.) A second worry for the formal semanticist who pursues (2) is that it may undermine her claim that semantics deals with complete propositions or truth-evaluable items. For instance, take the situation where the book Jack demonstrates is better than the book Jill is reading, but worse than the book she wrote. To assess Jack’s opening utterance as true or false in this situation, it might seem that we need first to determine what relationship Jack intended by his utterance of “Jill’s book.” Without this pragmatically determined aspect of what is said, it is argued, the sentence Jack

produces is simply not truth evaluable, it doesn’t express a complete proposition (cf. Recanati, 2003).

Clearly, then, although the formal semanticist can respond to the challenge of UCs with moves like (1) and (2) it is not at all obvious that these responses are sufficient. However, we should also be aware that there are problems to be faced on the other side, by the proponent of UCs, as well. One major worry is how to preserve our intuitive distinction between semantically relevant elements and elements which are only pragmatically relevant. Given the traditional conception of formal semantics, the answer to this question was clear: something is semantically relevant if it can be traced to the syntax of the sentence, it is pragmatically relevant otherwise. Once we admit of semantically relevant but syntactically unmarked elements, however, this way of characterizing the distinction is clearly unavailable, so the proponent of UCs owes us some other account. If Jack utters “I’ve eaten” then it might seem plausible to claim that the semantic content of this utterance is that he has eaten breakfast today, but he might also succeed in conveying the messages that he has eaten a cooked breakfast within the last hour, or that he is not hungry, or any number of further propositions. How do we decide which elements come to figure in the supposed semantic content of the utterance, and on what basis do we rule some conveyed messages as merely pragmatic?

A further worry seems to be that, for each additional contextual contribution we introduce, this contribution is itself open to further qualification: an utterance of “That banana is green” might intuitively be contextually enriched to that banana is green on its skin, but why stop here? Does the speaker mean to say that that banana is green all over its skin, that that banana is mostly green on a particular patch of skin, or that that banana is a bit green on this bit of surface skin and to a depth of degree n through the skin? The problem is that, for any piece of contextual information we introduce via a UC, this piece of information is itself likely to allow for a number of different contextual qualifications or sharpenings, and we seem to lack any principled reason to disallow these further sharpenings from appearing as part of the semantic content of the utterance. Yet, without such a reason, we run the risk of being launched on a slippery slope whereby the meaning of every utterance turns out to be a proposition which is somehow ‘complete’ in every respect. Yet no such conclusion seems palatable: it undermines notions of systematicity for meaning and makes it completely unclear how we ever come to learn and use a language given our finite cognitive resources (*see Compositionality: Semantic Aspects; Compositionality: Philosophical Aspects; Systematicity*).

So, it seems there are problems on both sides of the debate here and the question of how to handle pragmatic determinants of what is said remains a vexed one. The advocate of formal semantics needs to deliver an account both of overtly context-sensitive expressions – the indexicals, demonstratives, and tense markers of a natural language – and of the more covert kinds of context sensitivity discussed here. Reflecting on the multitude of ways in which a context of utterance can apparently affect issues of linguistic meaning certainly seems to show that the formal semanticist runs the risk of undervaluing the role played by pragmatic determinants in what is said. Yet it remains to be seen whether this constitutes merely a potential oversight on her behalf (cf. Stanley, 2000; Stanley and Szabo, 2000; Borg, 2004) or a fundamental failing of the whole approach (cf. Sperber and Wilson, 1986; Carston, 2002).

See also: Character versus Content; Compositionality: Philosophical Aspects; Compositionality: Semantic Aspects; Expression Meaning versus Utterance/Speaker Meaning; Semantics–Pragmatics Boundary; Systematicity; Truth Conditional Semantics and Meaning.

Bibliography

Bach K (1994). 'Semantic slack: what is said and more.' In Tsohatzidis S (ed.) *Foundations of speech act theory*:

philosophical and linguistic perspectives. London: Routledge. 267–291.

Bach K (1999). 'The semantics–pragmatics distinction: what it is and why it matters.' In Turner K (ed.) *The semantics/pragmatics interface from different points of view*. Oxford/New York: Elsevier Science. 65–83.

Borg E (2004). *Minimal semantics*. Oxford: Oxford University Press.

Cappelen H & Lepore E (2005). *Insensitive semantics: a defense of semantic minimalism and speech act pluralism*. Oxford: Blackwell.

Carston R (2002). *Thoughts and utterances*. Oxford: Blackwell.

Elugardo R & Stainton R (2004). 'Shorthand, syntactic ellipsis and the pragmatic determinants of what is said.' *Mind and Language* 19(4), 359–471.

Perry J (1986). 'Thought without representation.' *Proceedings of the Aristotelian Society Supplementary Volume* 40, 263–283.

Recanati F (2002). 'Unarticulated constituents.' *Linguistics and Philosophy* 25, 299–345.

Recanati F (2003). *Literal meaning*. Cambridge: Cambridge University Press.

Sperber D & Wilson D (1986). *Relevance: communication and cognition*. Oxford: Blackwell.

Stanley J (2000). 'Context and logical form.' *Linguistics and Philosophy* 23, 391–434.

Stanley J & Szabo Z (2000). 'On quantifier domain restriction.' *Mind and Language* 15, 219–261.

Taylor K (2001). 'Sex, breakfast, and descriptus interruptus.' *Synthese* 128, 45–61.

Predication

S Rothstein, Bar-Ilan University, Ramat Gan, Israel

© 2006 Elsevier Ltd. All rights reserved.

The predication relation was introduced into discourse about language by Aristotle in 'On Interpretation,' (paragraphs 5 and 6) when he defined a proposition as an instance of predication, which "asserts or denies something of something." He thus introduced into philosophical and linguistic discussion two central ideas: (a) the syntactic idea that a proposition has a binary structure, with one element, the subject, making reference to an entity and the other expressing a property, and (b) the semantic idea that asserting a proposition involves an act of asserting that an object, the reference of the subject argument, has a property expressed by the predicate. Thus, an assertion of *John is happy* is true if the denotation of the subject argument *John* has the property expressed by the predicate *is happy*.

The concept of predication played a central role in traditional grammatical discourse, and the division of a sentence into subject and predicate, became a commonplace of traditional parsing techniques. However, an understanding of what a subject and predicate actually might be waited until the techniques of modern logic and linguistics, when two questions in particular came to the center of the discussion. The first is the question of why the predication relation is so important, which could be discussed only after Frege had formulated his distinction between saturated and unsaturated linguistic entities. The second question is why one referential argument, the subject, has a privileged or distinguished position in the sentence, and progress on this point had to wait for the structural tools of modern generative syntax. This issue in fact never arose for Aristotle, since he dealt only with single-argument sentences such as *The ship sailed*, but it arises as soon as we look at sentences such as *The sailors sailed the ship*.

The major breakthrough for predication (and in fact semantics in general) came with the work of Frege at the end of the 19th century. Frege drew the distinction between saturated and unsaturated expressions. A formally unsaturated expression is an expression that contains an empty position filled by a variable, as in the mathematical expression $x + 2$, whereas a saturated expression is one that we naturally think of as denoting an object (for Frege, either an individual or a truth value). An unsaturated expression, or function, applies to a saturated expression or argument and yields a value or an output. Thus $x + 2$ applies to an argument, for example 3, to yield a value that is the denotation of the saturated expression. In this case, the value of the saturated expression $3 + 2$ is the same as the denotation of 5.

Frege saw that linguistic expressions could be analyzed in the same way. The expression *The capital city of x* can be analyzed as a function that has as its denotation, or value, a different city depending on which name replaces the variable x . *The capital city of Germany* denotes Berlin and *The capital city of the UK* denotes London. Similarly, a predicate such as *x is happy* can be applied to an argument expression such as *John* or *the boy* to yield a sentence, which has the value true or false, depending on whether or not John is or is not happy.

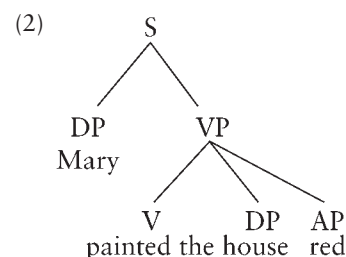
Frege's insight that complex expressions could be built up compositionally from simple saturated and unsaturated expressions, by a recursive process of functional application, was at the root of much progress in logic and semantics, as can be seen in the work of Montague (1970) and the formal semanticists who followed, and the work of Bar-Hillel and the categorical grammarians, all of whom focused on issues of compositionality. The issue of predication entered the syntactic discourse through the seminal work of Halliday (1967), who drew attention to secondary predicates, which were predicated of and expressed properties of local subjects, rather than sentential subjects. Halliday (1967) brought to the attention of linguists that predicative adjectives occurred with quite a limited range of meanings in constructions such as (1), and that the relation between the predicate and its local subject (in bold and italics, respectively) mirrored the semantic relation between the predicate of a sentence and the sentential subject, in the sense that if the sentence asserted was true, the subject had to have the property expressed by the predicate.

- (1a) Mary painted the *house* **red**.
- (1b) *John* drove the car **drunk**.
- (1c) John drove *the car* **broken**.

Example (1a) is an instance of resultative predication, with the AP predicate expressing a property that the subject is asserted to have at the end of the action

expressed by the verbal predicate; (1b) and (1c) are examples of depictive predication, with the AP predicate expressing a property that the subject is asserted to have while the action expressed by the verbal predicate is going on. Resultatives are typically predicated of direct object arguments, while depictives, as the examples in (1) show, may be predicated of subject or direct object arguments.

It was Edwin Williams, who in a series of papers (Williams, 1980, 1983, 1987) brought predication into the domain of generative syntax. Although there had been quite a lot of discussion around the notion of subject, Williams was the first to tie it into a discussion of the subject predicate relation. For him, the central questions were how to explain the syntactic relation of a secondary predicate to its subject (in more modern theoretical language, to explain how the predicate AP is licensed, or integrated into the structure of the sentence), how to explain why the subject argument is distinguished, and how to draw the appropriate structural parallels between secondary and main clause predicates. He argued that predication is essentially a thematic relation, and that the syntactic predication relation was expressed via the assignment of thematic roles. A lexical head assigns thematic roles to its thematic arguments, and in general all but one of the thematic arguments is assigned to an argument syntactically realized inside the projection of the thematic head. The thematic role assigned outside the projection of the head is external. Williams argued that the subject of a predicate is the argument assigned an external argument, and the primacy of the subject follows from the basic syntactic distinction between external and internal arguments. Thus, in a sentence such as (1a), assigned the structure in (2), the adjective *red* assigns a thematic role outside the maximal AP projection that it heads to *the house* and that therefore is its subject. Williams showed that this mechanism of external theta-role assignment could be used to express sentential subject predicate relations also. In (1a), *Mary* is assigned a thematic role 'agent' by *paint* outside the VP projection of the head, and is thus the subject of the whole predicate *paint the house red*.



What Williams captured precisely in this account is the duality of the relation between the sentential subject and the verb: the subject is a thematic argument

of the verb, but subject of the whole predicate phrase that the verb heads. Williams was not explicit about the semantic interpretation of the predication relation, but he assumes that the adicity of the lexical head (the number of thematic roles it assigns) directly reflects the adicity of a semantic relation denoted by the lexical head and that thematic role assignment is a syntactic reflection of a semantic relation. Williams was thus making implicit use of Frege's classification of linguistic objects into saturated and unsaturated. This became explicit in Williams (1987), where he presents an account of the mapping between syntactic representation and semantic representation, where he argues that unassigned thematic roles are interpreted as unsaturated variables in the semantic representation. Furthermore, argument traces (NP, or rather DP, traces) are to be analyzed as variables in the same way, and thus NP movement is ultimately reducible to the predication relation.

Rothstein (1983) and later Rothstein (1995, 2001) followed Williams in identifying predication as a central mechanism in syntactic theory, but argued against identifying syntactic unsaturatedness with unassigned thematic roles. She argued that in order to analyze a sentence generally as an instance of a predication relation, it is necessary to recognize cases of predication that do not involve a thematic relation between the head of the predicate and the subject. This is because of the existence of pleonastics, or non-thematic subjects, examples of which are given in (3):

- (3a) It seems that John will not arrive on time.
- (3b) Il a été mangé trois pommes. (French)
it has been eaten three apples.
'Three apples have been eaten.'
- (3c) ze meatzben Se dani meaxer. (Hebrew)
it irritates that dani is late.
'It is irritating that Dani is late.'

Rothstein argued that the Fregean saturated/unsaturated distinction is a primitive typal syntactic distinction between two different kinds of syntactic constituents: arguments and predicates. The typically referential constituents, DP and CP, are syntactically saturated and canonically serve as arguments, while the AP, PP, VP, and NP constituents are inherently unsaturated and can serve as predicates. Unsaturated constituents are typically saturated either directly or indirectly by predication (although NP arguments are saturated via determiner binding when they appear in argument position). Crucially, the syntactic relation between the pleonastic subject *it* in (3a) and the VP *seems that John will not arrive on time* is exactly the same predication relation as holds between the subject *John* and predicate *is happy* in the sentence *John is happy*. Rothstein argued that under this analysis,

the distribution of pleonastic subjects in English could be completely predicted, and that the obligatoriness of the subject (expressed in Chomsky's 1982 Extended Projection Principle) could be explained. In particular, she gave an account of the obligatoriness of the pleonastic in complement position of verbs such as *consider*, illustrated in (4a), in which she argues that pleonastics appear as the subject of small clause predications, and of the optional movement of CP direct objects to subject position illustrated in (4b):

- (4a) I consider it ridiculous that John is late again.
- (4b) It has long been believed that the earth is flat.

In this approach, thematic subjects are just a proper subset of subjects, and thematic role assignment to subject position uses the predication relation, just as thematic role assignment to internal arguments is via Case assignment relations or prepositions. Rothstein argued that there were two different kinds of predication: primary, or clausal, predication and secondary, or adjunct, predication. The particular constraints on adjunct predication (see Rothstein, 2003) mean that secondary predicates are essentially thematic, while primary predication need not be so.

Another approach to predication, which starts from different assumptions, but has some of the same syntactic consequences as Rothstein's, is Gennaro Chierchia's. Working within the framework of property theory, he proposed treating semantic properties as having two modes of presentation, an unsaturated predicate and its individual property correlate, π . He is concerned primarily with semantic issues, in particular, how properties can be predicated of themselves, as in *Being wise is wise* or *Cheerfulness is cheerful*. In Chierchia's theory, a VP denotes a saturated property, and Inflection denotes a function from individual properties into propositional functions, which applies to a VP and yields a propositional function, independently of whether the VP requires a thematic subject. Chierchia differs from Rothstein, who claims that the VP is formally a predicate, and he does not extend his theory to deal with noninflected predicates, but both approaches have in common that the clause is structurally constrained to be an instance of predication, independent of the thematic properties of the predicate or its head. Positing a structural, nonthematic predication relation has been quite fruitful in terms of syntactic explanatoriness. The structural approach to predication has been quite productive in leading to analyses of propositional structure, in particular with respect to the distribution of pleonastics and nonthematic predicates and copula constructions.

A second direction on which predication studies have focused has been the variety and differing

properties of secondary predicates crosslinguistically. As mentioned above, there seem to be two kinds of secondary predicates, resultatives and depictives, where resultatives are generally restricted to taking direct objects as their subjects, while depictives can apply to both subjects and objects. Much work has focused on the particular syntactic properties associated with different secondary predicates in different languages, for example with the structural constraints on subject predicate relations (Demonte, 1986), Case properties associated with different secondary predicates (especially in Russian, e.g., Filip, 2001), whether secondary predicates are independent predicates or form a lexical complex predicate with the verb in Germanic (Neeleman, 1994; Neelman and Weerman, 1993; Wunderlich, 1997), whether or not resultatives are truly oriented toward the direct object, (Rappaport Hovav and Levin, 2001; Wechsler, 1997; Kim, 1993), and how best to capture the semantic constraints on their interpretations. Some of the most interesting syntactic work in recent years has emerged from the attempts to give a unified semantics for depictive and resultative secondary predicates (see, e.g., Rothstein, 2003) and work by Cormack and Smith (1994, 1999), who suggest that secondary predication should be seen as a kind of serial verb construction for languages that do not otherwise allow them. While this idea has not yet been worked out in any detail, it seems a potentially fruitful direction in which to take the syntax of predication.

Bibliography

- Aristotle (1941). *The basic works*. In McKeon R (ed.). Random House.
- Chierchia G (1984). *Topics in the syntax and semantics of infinitives and gerunds*. Ph.D. diss. University of Massachusetts, Amherst.
- Chierchia G (1986). 'Predication.' *Linguistic Inquiry* 17, 417–443.
- Collins C (1997). 'Argument sharing in serial verb constructions.' *Linguistic Inquiry* 28, 461–497.
- Cormack A & Smith N (1994). 'Serial verbs.' In *Working Papers in Linguistics* 6. *Phonetics and Linguistics*. London: University College London. 63–88.
- Cormack A & Smith N (1999). 'Why are depictives different from resultatives?' In Iten C & Neelman A (eds.) *Working Papers in Linguistics* 11. *Phonetics and Linguistics*. London: University College London. 251–284.
- Demjjanoff A & Strigin A (2003). 'Real adjuncts in instrumental in Russian.' In Lang E, Maienborn C & Fabricius-Hansen C (eds.) *Modifying adjuncts*. Berlin: Mouton.
- Demonte V (1986). 'Predication and passive.' In Bordelais I, Contreras H & Zagona K (eds.) *Generative studies in Spanish syntax*. Dordrecht: Foris. 51–66.
- Filip H (2001). 'The semantics of case in Russian secondary predicates.' *Proceedings of SALT XI*. 191–211.
- Franks S & Hornstein N (1992). 'Secondary predicates in Russian and proper government of PRO.' In Higginbotham J, Larson R, Iatridou S & Lahiri U (eds.) *Control and grammar*. Dordrecht: Kluwer.
- Frege G (1891). 'Function and concept.' In Geach P & Black M (eds.) *Translations from the Philosophical Writings of G Frege*, 1980 edn. NJ: Rowman and Littlefield. 21–41.
- Frege G (1892). 'Concept and object.' In Geach P & Black M (eds.) *Translations from the Philosophical Writings of G Frege*, 1980 edn. NJ: Rowman and Littlefield. 42–55.
- Halliday M (1967). 'Notes on transitivity and theme in English, Part I.' *Journal of Linguistics* 3, 37–81.
- Kim J-B (1993). 'Syntax and semantics of Korean resultative constructions.' In Kuno S, Lee I, Whitman J, Maling J, Kang Y-S & Kim Y-J (eds.) *Harvard Studies in Korean Linguistics* V, 471–482.
- Landman F (2001). *Indefiniteness and the type of sets*. Blackwell: Oxford.
- Montague R (1970). Universal grammar. *Theoria* 36, 373–398. Reprinted in *Formal philosophy: selected papers of Richard Montague*, Thomason R H (ed) New Haven: Yale.
- Napoli D (1992). 'Secondary resultative predicates in Italian.' *Journal of Linguistics* 28, 53–90.
- Neeleman A (1994). *Complex predicates*. Ph.D. diss. Utrecht University.
- Neeleman A & Weerman F (1993). 'The balance between syntax and morphology: Dutch particles and resultatives.' *Natural Language and Linguistic Theory* 11, 433–475.
- Pesetsky D (1982). *Paths and categories*. Ph.D. diss. MIT.
- Rappaport Hovav M & Levin B (2001). 'An event-structure account of English resultatives.' *Language* 77, 766–797.
- Rothstein S (1983). *The syntactic forms of predication*. Ph.D. diss. MIT.
- Rothstein S (1995). 'Pleonastics and the interpretation of pronouns.' *Linguistic Inquiry* 26, 499–529.
- Rothstein S (2001). *Predicates and their subjects*. Dordrecht: Kluwer.
- Rothstein S (2003). 'Secondary predication and aspectual structure.' In Lang E, Maienborn C & Fabricius-Hansen C (eds.) *Modifying adjuncts*. Berlin: Mouton.
- Safir K (1983). 'On small clauses as constituents.' *Linguistic Inquiry* 14, 730–735.
- Schein B (1982/1995). 'Predication.' In Cardinaletti A & Guasti T (eds.) *Small clauses, Syntax and Semantics* 28, 49–76.
- Simpson J (1983). 'Resultatives.' In Levin L, Rappaport M & Zaenen A (eds.) *Papers in lexical-functional grammar*. Bloomington: Indiana University Linguistics Club.
- Takezawa K (1993). 'Secondary predication and locative/goal phrases.' In Hasegawa N (ed.) *Japanese syntax in comparative grammar*. Tokyo: Kurosio Publishers. 45–77.
- Washio R (1997). 'Resultatives, compositionality and language variation.' *Journal of East Asian Linguistics* 6, 1–49.
- Washio R (1999). 'Some comparative notes on resultatives.' In Muraki M & Iwamoto E (eds.) *Linguistics: in search of the human mind – A Festschrift for Kazukoe Inoue*. Tokyo: Kaitakusha.

- Wechsler S (1997). 'Resultative predicates and control.' *Texas Linguistics Forum* 38, 307–321.
- Wechsler S & Noh B (2001). 'On resultative predicates and clauses: parallels between Korean and English.' *Language Sciences* 23, 391–423.
- Williams E (1980). 'Predication.' *Linguistic Inquiry* 11, 203–238.

- Williams E (1983). 'Against small clauses.' *Linguistic Inquiry* 14, 287–308.
- Williams E (1987). 'NP trace and theta-theory.' *Linguistics and Philosophy* 10, 203–238.
- Wunderlich D (1997). 'Argument extension by lexical adjunction.' *Journal of Semantics* 14, 95–142.

Presupposition

P A M Seuren, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

© 2006 Elsevier Ltd. All rights reserved.

Introduction

A presupposition is a semantic property of a sentence making that sentence fit for use in certain contexts and unfit for use in others. This property is partly based on the fact that if a sentence *B* presupposes a sentence *A* ($B \gg A$), then *B* entails *A* ($B \models A$): whenever *B* is true, *A* is necessarily also true, given the same situation referred to, in virtue of the meanings of *B* and *A*. Presuppositions (*P-entailments*) are thus a subclass of entailments. Entailments that are not presuppositional are *classical* or *C-entailments*. (1) illustrates a C-entailment (\models_c); (2a–d) illustrate P-entailments (\gg):

- (1) Jack has been murdered. \models_c Jack is dead.
- (2a) Jack lives in Manchester. \gg Jack exists.
- (2b) Jill has forgotten that Jack is her student. \gg Jack is Jill's student.
- (2c) Jack is divorced. \gg Jack was once married.
- (2d) Only Jack left. \gg Jack left.

(2a) illustrates *existential presupposition*. (2b) exemplifies *factive presuppositions* (Kiparsky and Kiparsky, 1971): the factive predicate (*have forgotten*) requires the truth of the *that*-clause. (2c) is a case of *categorical presupposition*, derived from the lexical meaning of the main predicate (*be divorced*). (2d) belongs to a *remainder category*, the presupposition in question being due to the particle *only*.

There are various differences between P-entailments and C-entailments. When $B \gg A$, *A* is somehow 'prior' to *B*, restricting the domain within which *B* is interpretable. Presuppositions present themselves specifically, whereas C-entailments are 'unguided' and thus lack the function of restricting the interpretation domain. This makes presupposition relevant for the cognitive aspects of linguistic information transfer.

There is also a logical difference, especially regarding *negation*, the central operator in any logic.

In standard logic a sentence that is false for any reason whatsoever is made true by the preposed negation. In language, however, negation is sensitive, in default uses, only to C-entailments, leaving the P-entailments intact. Suppose (2d) is false on account of a C-entailment's falsity, for example because other people left as well. Then (3a), the negation of (2d), is true and (3b) is coherent, as expected. Not so when a P-entailment of (2d) is false, as in (3c), where the presupposition 'Jack left' is denied. (3c) is incoherent because *Not only Jack left* still presupposes that Jack left ('!!' indicates incoherence):

- (3a) Not only Jack left.
- (3b) Not only Jack left: other people left as well.
- (3c) !!Not only Jack left: Jack did not leave.

This raises the question of the truth value of (3a) in cases where, say, Jack did not leave. In standard logic, the entailment from both (2d) and its negation (3a) to *Jack left* means that if Jack did not leave, both (3a) and (2d) are false simultaneously, which violates the Principle of Contradiction ('a sentence and its negation cannot be true or false simultaneously'). Standard logic thus rules out falsity for *Jack left*, because its falsity leads to a contradiction. This makes *Jack left* a necessary truth. But *Jack left* is, if true, contingently so. Therefore, standard logic is inadequate for presuppositions.

Many feel that this calls for a rejection of the Principle of the Excluded Third or PET ('any sentence expressing a proposition is either true or false, without any values in between or outside') and for the introduction of a *third truth value* into the logic of language. Others prefer to keep standard bivalent logic intact and seek a *pragmatic* way out in terms of conditions of use. This question is discussed in the next section.

Operational Criteria

Presuppositions are detectable ('observable') irrespective of actual token use. Like C-entailments, P-entailments can be read off isolated sentences,

regardless of special context. Yet they evoke a context. (2a) requires it to be contextually given that there exists someone called 'Jack' and thus evokes such a context, asserting that he lives in Manchester; (2b) evokes a context where Jack is Jill's student, asserting that Jill has forgotten that; (2c) requires a context where Jack was once married, and asserts that the marriage has been dissolved; and (2d) requires a context where Jack left, while asserting that no one else left. This, together with the entailment criterion, provides a set of operational criteria to recognize presuppositions.

First, if $B \gg A$ then $B \models A$. The usual heuristic criterion for an entailment $B \models A$ is the incoherence of the juxtaposition of *not*(A) with B. On the whole, this suffices as a criterion. For example, (4a) does not entail, and therefore does not presuppose, (4b), because (4c) is still coherent.

- (4a) Lady Fortune neighs.
- (4b) Lady Fortune is a horse.
- (4c) Lady Fortune is not a horse, yet she neighs.

But this criterion overkills when the entailing sentence is qualified by an epistemic possibility operator, like English *may*, as in (5a), which does not entail (5b), even though (5c) is incoherent. Epistemic possibility requires compatibility of what is said to be possible with what is given in discourse or knowledge. Therefore, if $B \models A$, then with *not*(A) in the knowledge base, *possibly*(B) results in inconsistency, although *possibly*(B) does not entail A. The entailment criterion can be refined, without loss of generality, by testing the (in)coherence of the juxtaposition of *possibly* (*not*(A)) with B, as in (5d). Because (5d) is coherent, (5a) \models (5b):

- (5a) Jack may have been murdered.
- (5b) Jack is dead.
- (5c) !! Jack is not dead, yet he may have been murdered.
- (5d) Jack may not be dead, yet he may have been murdered (and thus be dead).

To distinguish P-entailments from C-entailments further criteria are needed. First there is the *projection criterion*: if $B \gg A$ and B stands under an entailment-canceling operator like *possibly* or *not* or *believe*, A survives not as a P-entailment but as a more or less strongly invited presuppositional inference ($>$). Generally, $O(B_A) > A$, where ' B_A ' stands for ' B presupposing A' and 'O' for an entailment-canceling operator. In standard terminology, the presupposition A is *projected* through the operator O. The conditions under which presuppositions of embedded clauses are projected through higher operators constitute the *projection problem of presupposition*.

Projection is typical of P-entailments, as in (6), not of C-entailments, as in (7):

- (6) Jill believes that Jack is divorced $>$ Jack was once married.
- (7) Jill believes that Jack has been murdered \nrightarrow Jack is dead.

The projection criterion is mostly used with negation as the entailment-canceling operator. Strawson (1950, 1952) held, incorrectly, that presupposition is always preserved as entailment under negation. In his view, a sentence like:

- (8) The present king of France is not wise.

still presupposes, and thus entails, that there exists a king of France, who therefore, if (8) is true, must lack wisdom. Although presupposition is, in fact, normally weakened to invited inference under negation, Strawson's 'negation test' became the standard test for presupposition. Provided the condition of 'entailment' is replaced by that of 'at least invited inference,' the test is sound.

Then there is the *discourse criterion*: a discourse bit A *and/but* B_A (with allowance for anaphoric processes) is felt to be orderly and well planned – that is, *sequential*. The condition of sequentiality is used to characterize stretches of acceptable text that have their presuppositions spelled out (' \checkmark ' signals sequentiality):

- (9a) \checkmark There exists someone called 'Jack,' and he lives in Manchester.
- (9b) \checkmark Jack is Jill's student, but she has forgotten that he is.
- (9c) \checkmark Jack was married, but he is divorced.
- (9d) \checkmark Jack left, and he is the only one who did.

C-entailments and inductive inferences behave differently. When they precede their carrier sentence the result may still be acceptable, yet there is a qualitative difference, as shown in (10a,b), where a colon after the first conjunct is more natural ('#' signals nonsequential but coherent discourse):

- (10a) #Jack is dead: he has been murdered.
- (10b) #Jack earns money: he has a job now.

The discourse criterion still applies through projection: A *and/but* O(B_A) is again sequential (the entailment-canceling operators are printed in italics):

- (11a) \checkmark Jack really exists, and Jill *believes* that he lives in Manchester.
- (11b) \checkmark Jack is Jill's student, but she has *probably* forgotten that he is.
- (11c) \checkmark Jack was once married, and he is *not* divorced.
- (11d) \checkmark Jack left, and he is *not* the only one who did.

These tests reliably set off P-entailments from C-entailments.

The Logical Problem

The Threat to Bivalence

The first to see the threat posed by presuppositions to standard logic was Aristotle's contemporary Eubulides of Miletus (Kneale and Kneale, 1962: 113–117). He formulated (besides other paradoxes such as the Liar) the paradox of the *Horned Man* (Kneale and Kneale, 1962: 114): “What you have not lost you still have. But you have not lost your horns. So you still have horns.” This paradox rests on presupposition. Read *B* for *You have lost your horns* and *A* for *You had horns*. Now $B \gg A$ (the predicate *have lost* induces the presupposition that what has been lost was once possessed).

Eubulides implicitly assumed that P-entailments are preserved under negation: $B \models A$ and $\text{not}(B) \models A$. Under PET, this would make *A* a logically necessary truth, which is absurd for a contingent sentence like *You had horns*. To avoid this, PET would have to be dropped, very much against Aristotle's wish. Although Aristotle himself was unable to show Eubulides wrong, there is a flaw in the paradox. It lies in the incorrectly assumed entailment in the first premise “What you have not lost you still have.” For it is possible that a person has not lost something precisely because he never had it.

The same problem was raised by Strawson (1950, 1952), but with regard to existential presuppositions. Like Eubulides, Strawson assumed full entailment of presupposition under negation and concluded that PET had to go. For him, nonfulfillment of a presupposition leads to both the carrier sentence and its negation lacking a truth value altogether. Frege (1892) had come to the same conclusion, though from a different angle. In a sentence like:

(12) The unicorn ran.

analyzed as ‘Run(the unicorn)’, the subject term lacks a referent in the actual world, though the existence of such a referent is presupposed. That makes it impossible to test the truth of (12): since there is no unicorn, there is no way to check whether it actually ran. Therefore, Frege (and Strawson) concluded, (12) lacks a truth value.

This posed a profound problem for standard logic in that the applicability of standard logic to English would have to be made dependent on contingent conditions of existence – a restriction no logician will accept. In the effort to solve this problem two traditions developed, the *Russell tradition* and the *Frege–Strawson tradition*.

The Russell Tradition

In his famous 1905 article, Russell proposed a new analysis for sentences with definite terms, like (13a).

Putting the new theory of quantification to use, he analyzed (13a) as (13b), or ‘there is an individual *x* such that *x* is now king of France and *x* is bald, and for all individuals *y*, if *y* is now king of France, *y* is identical with *x*’:

- (13a) The present king of France is bald.
(13b) $\exists x [\text{KoF}(x) \wedge \text{Bald}(x) \wedge \forall y [\text{KoF}(y) \rightarrow x = y]]$

In order to save bivalence, Russell thus replaced the time-honored subject-predicate analysis with an analysis in which the definite description *the present king of France* no longer forms a constituent of the logically analyzed sentence, but is dissolved into quantifiers and propositional functions.

The negation of (13a) should be (13b) preceded by the negation operator, i.e. (14a). However, Russell held, speakers often prefer, for reasons best known to themselves, to interpret *The present king of France is not bald* as (14b), with internal negation over ‘Bald(*x*)’:

- (14a) $\neg \exists x [\text{KoF}(x) \wedge \text{Bald}(x) \wedge \forall y [\text{KoF}(y) \rightarrow x = y]]$
(14b) $\exists x [\text{KoF}(x) \wedge \neg \text{Bald}(x) \wedge \forall y [\text{KoF}(y) \rightarrow x = y]]$

This makes sentences like (8) ambiguous.

This analysis, known as Russell's *Theory of Descriptions*, was quickly accepted by logicians and philosophers of language, as it saved PET. At the same time, however, it drove logicians and linguists apart, as it defies any notion of sentence structure. Moreover, the ‘uniqueness clause’ in (13b), $\forall y [\text{KoF}(y) \rightarrow x = y]$, saying that only one king of France exists, is meant to account for the uniqueness expressed by the definite article. In fact, however, the definite article implies no claim to uniqueness of existence, only to discourse-bound uniqueness of reference. Then, this analysis is limited to definite descriptions and is unable to account for other kinds of presupposition. Factive and categorial presuppositions, and those derived from words like *all*, *still*, or *only*, fall outside its coverage.

An important objection is also that negation can only cancel presuppositions when it is a separate word (not a bound morpheme) and in construction with a finite verb. In all other cases, the negation fully preserves P-entailments. Thus, (3a), with *not* in construction with *only*, preserves the presupposition induced by *only*. Moreover, sentence-initial factive *that*-clauses preserve presuppositions even though the negation is constructed with the finite verb:

- (15a) That Jack left surprised Jill \gg Jack left.
(15b) That Jack left did not surprise Jill \gg Jack left.

Likewise for cleft and pseudocleft sentences:

- (16a) It was Jack who left / The one who left was Jack \gg Someone left.
(16b) It wasn't Jack who left / The one who left wasn't Jack \gg Someone left.

When cases like these, overlooked by the authors discussed, are taken into account and the logic is kept bivalent, the presuppositions of sentences like (2d) and (3a), (15a,b), or (16a,b) would again have to be necessary truths.

The same goes for:

(17a) All men are mortal \gg There exist men.

(17b) Not all men are mortal \gg There exist men.

In standard Predicate Calculus, however, (17a) does not entail (and thus cannot presuppose) that there exist men, whereas (17b) does, because ‘not all F is G’ is considered equivalent with ‘some F is not G,’ which entails the existence of at least one F. Yet both (17a) and (17b) satisfy the operational criteria given earlier. Standard Predicate Calculus thus seems to fit the presuppositional facts badly.

To account for other than existential presuppositions some have proposed to change Russell’s analysis into:

(18) $\exists x [KoF(x)] \wedge Bald(he)$

or ‘there is a king of France, and he is bald’. *He* is now no longer a bound variable but an anaphoric pronoun. With a logical mechanism for such anaphora (as in Kamp, 1981; Groenendijk and Stokhof, 1991), this analysis can be generalized to all categories of presupposition. A sentence B_A is now analyzed as A and B_A , and $Not(B_A)$, though normally analyzed as A and $Not(B_A)$ with small scope *not*, can also, forced by discourse conditions, be analyzed as $Not(A$ and $B_A)$, with large scope *not*. This analysis, which saves PET, is known as the *Conjunction Analysis* for presupposition.

Anaphora is needed anyway, because Russell’s analysis fails for cases like (19), where quantifier binding is impossible for *it*, which is in the scope of *I hope*, whereas *I hope* is outside the scope of *I know*:

(19) I know that there is a dog and I hope that it is white.

The Conjunction Analysis, however, still cannot account for the fact that (20a) is coherent but (20b) is not:

(20a) There is a dog and it is white, and there is a dog and it is not white.

(20b) !!There is a dog and it is white and it is not white.

(20a) speaks of two dogs, due to the repetition of *there is a dog*, but (20b) speaks of only one. Yet the Conjunction Analysis cannot make that difference, because the repetition of *there is a dog* makes no logical or semantic difference for it. Attempts have been made to incorporate this difference into the logic

(e.g., Kamp, 1981; Heim, 1982; Groenendijk and Stokhof, 1991) by attaching a memory store to the model theory that keeps track of the elements that have so far been introduced existentially.

Even then, however, the Conjunction Analysis still postulates existence for term referents whose existence is denied:

(21) Santa Claus does not exist.

The Frege-Strawson Tradition

Strawson (1950, 1952) was the first to oppose the Russell tradition. He reinstated the traditional subject-predicate analysis and discussed only existential presuppositions. Negation is considered presupposition-preserving. Sentences with presupposition failure are considered truth-valueless. Strawson’s definition of presupposition is strictly logical: $B \gg A =_{Def} B \models A$ and $Not(B) \models A$. This analysis requires a *gapped bivalent propositional calculus* (GBPC), shown in Figure 1.

Insofar as truth values are assigned, GBPC preserves standard logic. Moreover, * is ‘infectious’: when fed into a truth function it yields *. Remarkably, GBPC limits the applicability of logic to situations where the presuppositions of the sentences involved are true. The applicability of GBPC thus varies with contingent circumstances.

Wilson (1975) and Boër and Lycan (1976) side with Russell and criticize Strawson, showing examples of presupposition-canceling under negation: (22a–c) are coherent, though they require emphatic, discourse-correcting accent on *not*:

(22a) The present king of France is NOT bald: there is no king of France!

(22b) Jill has NOT forgotten that Jack is her student: Jack isn’t her student!

(22c) Jack is NOT divorced: he never married!

For these authors, classical bivalent logic is adequate for language; P-entailments differ from C-entailments only pragmatically. There would be a point if (a) a pragmatic explanation were available, and (b) presuppositions were always canceled under negation. But neither condition is fulfilled.

		\wedge B			\vee B		
\sim A	A	T	F	*	T	F	*
F	T	T	F	*	T	T	*
T	F	F	F	*	T	F	*
*	*	*	*	*	*	*	*

Figure 1 Strawson’s gapped bivalent propositional calculus (GBPC). Key: \sim , presupposition-preserving negation; T, truth; F, falsity; *, unvalued.

In fact, the presupposition-canceling ‘echo’ negation NOT of (22a–c) is impossible for cases that preserve P-entailments under negation:

- (23a) !!NOT only Jack left: he didn’t leave!
 (23b) !!NOT all students protested: there weren’t any students!
 (23c) !!That Jack left did NOT surprise Jill: he didn’t leave!
 (23d) !! The one who left was NOT Jack: nobody left!

Likewise for the negation required with *negative polarity items* (NPIs) in assertive main clauses (NPIs are printed in italics):

- (24a) !!Jack does NOT *mind* that he is in trouble: he isn’t in trouble!
 (24b) !!Jack has NOT come back *yet*: he never went away!
 (24c) !!Jill has NOT seen Jack *in weeks*: she doesn’t exist!

This analysis is thus fatally flawed.

The Trivalent Solution

One may envisage a three-valued logic, identical to standard bivalent logic but for a distinction between two kinds of falsity, each turned into truth by a separate negation operator. Minimal falsity (F1) results when all P-entailments are true but not all C-entailments, radical falsity (F2) when one or more P-entailments are false. Correspondingly, *minimal negation* (\sim) turns F1 into truth (T) and T into F1, leaving F2 unaffected, whereas *radical negation* (\simeq) turns F2 into T and both T and F1 into F1.

In Kleene’s (1938) *trivalent propositional calculus*, \wedge yields T only if both conjuncts are T, F1 when either conjunct is F1, and F2 otherwise. Analogously, \vee yields T when either conjunct is T, F1 only if both conjuncts are F1, and F2 otherwise. The corresponding tables are given in Figure 2 (where the value F2 is named ‘indefinite’ or I). This logic preserves all theorems of standard logic when bivalent \neg replaces trivalent \sim . Kleene’s calculus lacks the radical negation (\simeq), but comes to no harm if it is added.

Kleene’s calculus is used by some presuppositional logicians (e.g., Blau, 1978). It is empirically problematic in that it yields F1 for ‘ $A \wedge B$ ’ when either A or B is F2 whereas the other is F1, thus allowing presupposition failure in one conjunct while still considering

the conjunction as a whole free from presupposition failure. This makes no sense in view of *and* as a discourse incrementer. Kleene’s calculus is more suitable for vagueness phenomena with F2 as an umbrella value for all intermediate values between T and F (Seuren *et al.*, 2001).

In Seuren’s presuppositional propositional calculus TPC2 (Seuren, 1985, 2001: 333–383; Seuren *et al.*, 2001) the operators \wedge and \vee select, respectively, the highest and lowest of the component values ($F2 > F1 > T$), as shown in Figure 3. Classical negation (\neg), added for good measure, is the union of \sim and \simeq , but is taken not to occur in natural language, which has only \sim and \simeq . In TPC2, F2 for either conjunct yields F2 for ‘ $A \wedge B$ ’, as required.

TPC2 is likewise equivalent with standard bivalent logic under the operators \neg , \wedge , and \vee (Weijters, 1985). Thus, closed under (\neg , \wedge , \vee) standard bivalent logic is independent of the number of truth values employed, though any value ‘false’ beyond F1 will be vacuous. Moreover, in both generalizations with n truth values ($n \geq 2$), there is, for any value $i \geq 2$, a *specific negation* N^i turning i into T, values lower than i into F1, and leaving higher values unaffected. Thus, in TPC2, N^{F1} is \sim and N^{F2} is \simeq . Classical bivalent \neg is the union of all specific negations. Consequently, in the standard system, \neg is both the one specific negation allowed for and the union of all specific negations admitted. Standard logic is thus the most economical variety possible of a generalized calculus of either type.

The Discourse Approach

Presupposition is not defined, only restricted, by its logical properties:

- (25) If $B \gg A$, then $B \models A$ and $\sim B \models A$, and
 $\sim A / \simeq A \models \simeq B$.

(25) thus specifies necessary, but not sufficient, conditions for presupposition. Were one to adopt a purely logical definition, various paradoxical consequences would follow. For example, any arbitrary sentence would presuppose any necessary truth, which would make the notion of presupposition empirically vacuous.

			$\wedge B$			$\vee B$		
$\neg A$	$\sim A$	A	T	F1	F2	T	F1	F2
F1	F1	T	T	F1	F2	T	T	T
T	T	F1	F1	F1	F1	T	F1	F2
T	F2	F2	F2	F1	F2	T	F2	F2

Figure 2 Trivalent propositional calculus (TPC1).

				$\wedge B$			$\vee B$		
$\neg A$	$\simeq A$	$\sim A$	A	T	F1	F2	T	F1	F2
F1	F1	F1	T	T	F1	F2	T	T	T
T	F1	T	F1	F1	F1	F2	T	F1	F1
T	T	F2	F2	F2	F2	F2	T	F1	F2

Figure 3 Trivalent propositional calculus (TPC2).

Attempts have been made (Gazdar, 1979; Heim, 1982; Seuren, 1985, 2000) at viewing a presupposition A of a sentence B_A as restricting the interpretable use of B to contexts that admit of, or already contain, the information carried by A . Such an approach creates room for an account of the discourse-correcting 'echo' function of presupposition-canceling (radical) NOT. Horn (1985, 1989) correctly calls NOT *metalinguistic*, in that it says something about the *sentence* in its scope – though his generalization to other metalinguistic uses of negation is less certain. Neither TPC2 nor TPC1 can account for this metalinguistic property. This means that the argument of NOT is not a sentence but a quoted sentence. NOT('B_A') says about the sentence B_A that it cannot be sequentially incremented in a discourse refusing A .

Sequential incrementation to a discourse D restricts D to a progressively narrower section of the universe of all possible situations U , making the increment *informative*. Incrementation of A , or $i(A)$, to D restricts D to the intersection of the set of situations in which D is true and the set of situations where A is true. The set of situations in which a sentence or set of sentences X is true is the *valuation space* of X , or $/X/$. For D incremented with A we write ' $D + A$ '. $D + A$ is the conjunction of D and A , where D is the conjunction of all incremented sentences since the initial U . The *sequentiality condition* requires:

- (a) for any A , $/D/$ is larger than $/D + A/$
(informativity: remember that D is restricted by A);
- (b) if $B \gg A$ then $i(A)$ must precede $i(B)$ (not so when $B \models_c A$).

If A has not already been incremented prior to $i(BA)$ it is cognitively 'slipped in,' a process called *accommodation* or *post hoc suppletion*. A text requiring accommodation is not fully sequential, but still fully coherent.

On the assumption that D , as so far developed, is true, any subsequent sentence B must be valued T or F1, because F2 for B implies that some presupposition of B , and hence D as a whole, is not true. This assumption is made possible by the *Principle of Presumed Truth* (PPT), which says that it must be possible for any D to be true. The assumption that D is actually true blocks the processing of a new sentence that would be valued F2 in D . For example, let D contain $i(\sim A)$. Now B_A is blocked, because A is valued F1 (assuming that D is true). But NOT('B') can be incremented and is true under PPT, as it says about B that it cannot be incremented.

Therefore, $/D/$ must contain situations with sentences as objects. Cognitively speaking, this is perfectly

plausible, because speakers are aware of the fact that they utter words and sentences. That awareness enables them to refer back to words and sentences just uttered or expected to be uttered. Words and sentences as objects are a necessary corollary of any speech utterance. This corollary underlies the free mixing of object language and metalanguage in natural language. The prohibition issued by logicians against such mixing has no ground in natural language (Seuren, 2001: 125–130).

Natural language negation is, in a sense, ambiguous (depending on syntactic conditions) between presupposition-preserving (minimal) *not* and presupposition-canceling (radical) NOT. Many find this unacceptable, because ambiguities are normally language specific, whereas this ambiguity would appear to be universal. Yet the obvious question of what would be the overarching single meaning of the negation in all its uses has not, so far, been answered. Similar problems occur with other logical operators, especially with *and*, *or*, and *if*, as the following examples show:

- (26a) Do as I say and you will be a rich man.
- (26b) Don't come nearer, or you'll be a dead man.
- (26c) That's awful, or should I say 'dreadful'?
- (26d) Let's go home, or do you have a better idea?
- (26e) If you're tired, I have a spare bed.

In the case of *not* and the other logical operators, speech act factors as well as factors of metalinguistic use play an important role. Unfortunately, the grammar and semantics of both speech acts and metalinguistic use are still largely unexplored. Horn (1985) pointed out that English *not* is often used metalinguistically, as in:

- (27a) Not Lizzy, if you please, but the Queen is wearing a funny hat.
- (27b) She wasn't happy, she was ecstatic!

And he classifies radical NOT with the other metalinguistic cases. However, as pointed out in Seuren (2001: 345–347), NOT, though metalinguistic, differs from the other cases in that it can only occur in construction with the finite verb (the 'canonical position'), whereas the other metalinguistic negations can occupy any position normal *not* can occur in. (28a) is coherent, with a canonically placed NOT, but (28b,c), likewise with NOT, are incoherent, as NOT is in a non-canonical position:

- (28a) He did NOT only lose \$500. He only lost \$20.
- (28b) !!NOT only did he lose \$500. He only lost \$20.
- (28c) !!He NOT only lost \$500. He only lost \$20.

The question of the overall meaning description of the logical operators, in terms of which their strictly logical meaning would find a place, defines a research

project of considerable magnitude – a project that has so far not been undertaken in a coordinated way.

The Structural Source of Presuppositions

The source of at least three of the four types of presupposition distinguished earlier lies in the satisfaction conditions of the main predicate of the carrier sentence. The satisfaction conditions of an n -ary predicate P^n are the conditions that must be satisfied by any n -tuple of objects for P^n to yield truth. Thus, for the unary predicate *white* the conditions must specify when any object can truthfully be called 'white'. For the binary predicate *wash* they must specify when it can truthfully be said of any pair of objects $\langle i, j \rangle$ that '*i* washes *j*'.

A distinction is made between two kinds of lexical conditions, *preconditions* and *update conditions*. When a precondition is not fulfilled, the sentence is radically false; failure of an update condition yields minimal falsity. Fulfillment of all conditions gives truth.

The satisfaction conditions of a predicate P^n are specified according to the schema ($[[P^n]]$ is the extension of P^n):

- (29) $[[P^n]] = \{ \langle i^1, i^2, \dots, i^n \rangle : \dots (\text{preconditions}) \dots$
 $\quad \quad \quad | \dots (\text{update conditions}) \dots \}$
 or: 'the extension of P^n is the set of all n -tuples of objects $\langle i^1, i^2, \dots, i^n \rangle$ such that
 $\dots (\text{preconditions}) \dots$ and $\dots (\text{update conditions}) \dots$ '.

The satisfaction conditions of the predicate *bald*, for example, may be specified as follows (without claiming lexicographical adequacy):

- (30) $[[\text{Bald}]] = \{ i : i \text{ is normally covered, in}$
 $\quad \quad \quad \text{prototypical places, with hair, fur, or pile;}$
 $\quad \quad \quad \text{or } i \text{ is a tire and normally covered with tread } |$
 $\quad \quad \quad \text{the normal covering is absent} \}$

This caters for categorial presuppositions. Factive presuppositions are derived by the precondition that the factive clause must be true.

Existential presuppositions are derivable from the precondition that a specific term t of a predicate P^n refers to an object existing in the real world. P^n is then *extensional* with respect to t . *Talk about*, for example, is extensional with respect to its subject term, but not with respect to its object term, because one can talk about things that do not exist. The satisfaction conditions of *talk about* will thus be as in (31), where the asterisk on j indicates that *talk* is nonextensional with respect to its object term:

- (31) $[[\text{talk about}]] = \{ \langle i, j^* \rangle : \dots (\text{preconditions}) \dots$
 $\quad \quad \quad | \dots (\text{satisfaction conditions}) \dots \}$

The predicate *exist* lacks any preconditions and is to be specified as nonextensional with respect to its subject term:

- (32) $[[\text{exist}]] = \{ i^* \mid i \text{ is an object in the actual world} \}$

A definite subject of the verb *exist* must be represented somewhere in **D**, normally in some intensional subdomain, e.g., the subdomain of things that Jack keeps talking about, as in:

- (33) The man that Jack keeps talking about really exists.

The incremental effect of (33) is that the representation of the thing that is said to exist is moved up to the truth domain of **D**. This analysis requires the assumption of *virtual objects*.

The remainder category of presuppositions, induced by words like *only* or *still*, or by contrastive accent or (pseudo)cleft constructions, looks as if it cannot be thus derived. The choice here is either to derive them by *ad hoc* rules or to adopt a syntactic analysis in terms of which of these words and accents figure as (abstract) predicates at the level of semantic representation taken as input to the incrementation procedure.

Bibliography

- Blau U (1978). *Die dreiwertige Logik der Sprache. Ihre Syntax, Semantik und Anwendung in der Sprachanalyse*. Berlin-New York: De Gruyter.
- Boër S & Lycan W (1976). *The myth of semantic presupposition*. Indiana University Linguistics Club.
- Frege G (1892). 'Ueber Sinn und Bedeutung.' *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50.
- Gazdar G (1979). *Pragmatics: implicature, presupposition, and logical form*. New York-San Francisco-London: Academic Press.
- Geach P T (1972). *Logic matters*. Oxford: Blackwell.
- Groenendijk J & Stokhof M (1991). 'Dynamic predicate logic.' *Linguistics and Philosophy* 14, 39–100.
- Heim I (1982). 'The semantics of definite and indefinite noun phrases.' Ph.D. diss., University of Massachusetts at Amherst.
- Horn L R (1985). 'Metalinguistic negation and pragmatic ambiguity.' *Language* 61, 121–174.
- Horn L R (1989). *A natural history of negation*. Chicago: University of Chicago Press.
- Kamp H (1981). 'A theory of truth and semantic representation.' In Groenendijk J, Janssen T & Stokhof M (eds.) *Formal methods in the study of language 1*. Amsterdam: Mathematisch Centrum. 277–322.
- Kiparsky P & Kiparsky C (1971). 'Fact.' In Steinberg D & Jakobovits L (eds.) *Semantics: an interdisciplinary reader in philosophy, linguistics, and psychology*. Cambridge: Cambridge University Press. 345–369.

- Kleene S (1938). 'On notation for ordinal numbers.' *Journal of Symbolic Logic* 3, 150–155.
- Kneale W & Kneale M (1962). *The development of logic*. Oxford: Oxford University Press.
- Russell B (1905). 'On denoting.' *Mind* 14, 479–493.
- Seuren P A M (1985). *Discourse semantics*. Oxford: Blackwell.
- Seuren P A M (2000). 'Presupposition, negation and trivalence.' *Journal of Linguistics* 36, 261–297.
- Seuren P A M (2001). *A view of language*. Oxford: Oxford University Press.
- Seuren P A M, Capretta V & Geuvers H (2001). 'The logic and mathematics of occasion sentences.' *Linguistics & Philosophy* 24, 531–595.
- Strawson P F (1950). 'On referring.' *Mind* 59, 320–344.
- Strawson P F (1952). *Introduction to logical theory*. London: Methuen.
- Weijters A (1985). 'Presuppositional propositional calculi.' Appendix to Seuren (1985).
- Wilson D (1975). *Presuppositions and non-truth-conditional semantics*. London-New York-San Francisco: Academic Press.

Principles and Parameters Framework of Generative Grammar

R Freidin, Princeton University, Princeton, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

The main task of linguistic theory “must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning” (Chomsky, 1965: 27–28). The Principles and Parameters (henceforth P&P) framework of generative grammar attempts to fulfill this goal by postulating a specific set of general mechanisms and principles of grammar that are assumed to be part of the innate knowledge a language learner brings to the task of language acquisition – more precisely, the acquisition of a grammar of a human language. To the extent that much of a speaker’s knowledge of his/her language can be derived from this system of mechanisms and principles, the need for language-particular rules of grammar diminishes – thus contributing to an explanation for the rapidity and uniformity of language acquisition. From this perspective, much of the diversity across languages can be viewed as the result of systematic differences in the way the mechanisms and principles apply to the grammars of various languages. These differences appear to fall within a restricted range of options, referred to as parameters. For example, some languages require an overt subject in declarative sentences (e.g., English and French), whereas others allow declarative sentences without overt subjects, presumably where a subject pronoun is deleted (e.g., Spanish and Italian). These options are also assumed to be determined by the language faculty, thus specified as part of Universal Grammar (UG). Within this framework, research on language variation is based on the

working hypothesis that much of the apparent diversity among languages can be reduced to instances of parametric variation. This hypothesis has generated a substantial amount of promising research in the area of comparative grammar (see Jaeggli and Safir, 1989; Freidin, 1991, 1992, 1996; Baker, 1995, 2001; Rizzi, 2000).

It is worth noting that modern comparative grammar has a much more ambitious and wide-ranging goal than its 19-century predecessor. Both are concerned with establishing an explanatory basis for the relationships between human languages. The neo-grammarians of the 19th century were focused primarily on relationships between languages and groups of languages in terms of common ancestry. Contemporary generative grammar is concerned with a theory of grammar that is postulated to be an innate component of the human mind/brain, a faculty that all humans share as part of their genetic identity and that guides the acquisition of language. Thus the theory of grammar establishes the relationship among all human languages – not only those that happen to be related by historical accident (e.g., common ancestry).

The P&P approach to the study of human language came into focus in the late 1970s and has evolved substantially over the past decade as basic assumptions and proposals have been subjected to a critical re-evaluation within the Minimalist Program (see Chomsky, 1995b; Freidin, 1997). Initially, the P&P framework was based on a system of grammatical principles, formulated in terms of fundamental notions of grammatical analysis (e.g., case, agreement, binding, and government). These principles formed the core of modules (e.g., Case and government) that operated within derivations to determine the behavior of syntactic rules, the mechanisms that generate syntactic representations for linguistic expressions. They applied either to derivations (i.e., the application of rules) or to the representations constructed

by derivations. Over the past decade, significant parts of this modular theory of grammar have been eliminated under minimalism. The conditions on rule application that remain are for the most part formulated in terms of principles of efficient computation, while conditions on representations have been recast in terms of economy principles that apply at the two external interfaces: Phonetic Form (PF) at the sensorimotor interface and Logical Form (LF) at the conceptual-intensional interface.

The system of grammatical principles initially proposed as the core of P&P evolved out of the study of conditions on the application of transformations, beginning with Chomsky's work on the A-over-A Principle in the early 1960s (see Chomsky, 1964) and continuing with Ross' work on island constraints (Ross, 1967, 1984) and Chomsky's extension of the conditions framework (Chomsky, 1973). One of the first important breakthroughs in modern comparative grammar came from the work of Kayne (1975) and Quicoli (1976a, 1976b), which demonstrated that certain abstract conditions on the application of transformations, postulated for the analysis of English in Chomsky (1973), also applied to some very different constructions in Romance involving clitics and quantifiers. Another crucial step in the development of the current approach was achieved with the demonstration (Chomsky, 1976) that under Trace Theory the behavior of transformational rules followed from general principles and therefore that transformations could be stated in an optimally general form – essentially as bare elementary transformational operations (e.g., 'substitute α ' or 'adjoin α to β '). Transformational rules in such general form could be viewed as rules of UG (the theory of grammar) rather than as language-specific rules of a particular grammar (or as construction specific rules [e.g., a relative clause transformation]).

The study of filters (conditions on representations) provides another major element of the initial framework (see Chomsky and Lasnik, 1977), one which raises important empirical questions regarding the status of various levels of syntactic representation. This line of research led to the realization of the central role of the notions of Case and government for the theory of grammar (due to Jean Roger Vergnaud, see Rouveret and Vergnaud, 1980), and ultimately to Chomsky's (1981) reformulation and extension of the entire theory, commonly referred to as government and binding (or GB) theory.

Chomsky (1981) related phrase structure theory and the theory of transformations in a new way. Up until this version of the theory of generative grammar, the theory of phrase structure and the theory of transformations had developed more or

less independently: phrase structure rules and transformations were considered independent grammatical mechanisms, both necessary for constructing syntactic representations. By defining a notion of government in terms of the theory of phrase structure and then establishing the major role it plays in the various conditions that determine the behavior of transformations, Chomsky identified a conceptual thread that links the two theories.

During the past decade under the minimalist program, the connection between phrase structure and transformations has been reconstructed so that it is only transformations that construct syntactic representations. The new bare phrase structure theory (see Chomsky, 1995a) radically revises the notion of syntactic derivation in a way that significantly affects the organization and function of grammar under the P&P framework.

The remainder of this article sketches some of the essential parts of the current P&P framework. For an account of the preminimalist version of this framework, see Freidin (1994a).

Mechanisms for Phrase Structure Representations

Under the current P&P framework, syntactic (i.e., phrase structure) representations are constructed by a single transformational operation Merge, an adjunction operation that concatenates two syntactic objects and labels the resulting syntactic object with the syntactic category features of one of the two elements concatenated. Syntactic derivations begin with an array of lexical items selected from the lexicon, a numeration N . Merge then operates on N , constructing a phrase marker containing all the elements in N . In a derivation, lexical insertion results when Merge applies to a lexical item in N . Thus, when Merge applies to a preposition *from* and a NP *my sons* (itself constructed from the merger of two lexical elements), the result is a PP *from my sons* containing a P and a NP as constituents, where *from*, the head of the construction, projects its syntactic category P to label the new construct. The operation is constrained by an Inclusiveness Condition (1), which radically restricts the kind of syntactic objects that can be created in a derivation.

- (1) Inclusiveness Condition: any structure formed by the computation (in particular, PF and LF) is constituted of elements already present in the lexical items selected for N ; no new objects are added in the course of computation apart from rearrangements of lexical properties.
(Chomsky, 1995b: 228)

Condition (1) prohibits the categorial distinction of levels of phrasal projection and the use of indices. Merge is further constrained under the assumption that it is a strictly binary operation and therefore there can only be two immediate constituents of any phrasal category, never more and never less.

Given Merge, every phrase in a phrase structure representation has a unique head, the element that projects its syntactic category features as the label of the phrase. The application of Merge is further constrained by the order in which a head is merged with other constituents. A complement (generally an argument of the head when the head is some form of predicate [e.g., verb, adjective, or derived nominal]) is always concatenated with the head before other kinds of constituents. A specifier is always concatenated with the phrasal projection of the head last and therefore is always an immediate constituent of a maximal phrasal projection. Adjuncts (generally modifiers) are concatenated with a head (when there are no complements) or a phrasal projection of the head after complements and before the specifier.

The application of Merge that concatenates two independent syntactic objects (lexical items from the lexicon or phrases constructed from the concatenation of lexical items) is called External Merge (EM), in contrast to the application of Merge that applies to a single syntactic phrase by adjoining a subphrase of it to the phrase itself (i.e., to its root). This latter application of Merge, which is called Internal Merge (IM), performs movement operations as in the case of simple passive sentences.

(2) Seymour was arrested in Shanghai.

Because the subject NP *Seymour* is interpreted at LF as the object of *arrested*, it is first merged (via EM) as the object of the verb and then moved (via IM) to subject position where it is pronounced (at PF). At LF, the syntactic representation of (1) contains a copy of subject in object position.

(3) Seymour_i was [_{VP} arrested Seymour_i in Shanghai]

(Although indices are prohibited under the Inclusiveness Condition [1], copies here are indicated here via coindexing as a notational convenience.) IM adjoins a copy of the object to the Tense Phrase (TP) *was arrested Seymour in Shanghai* and projects the label T, thereby creating the specifier position of TP.

IM involves only the adjunction operation, i.e. only a single elementary transformational operation. The deletion operation applies to unpronounced copies of syntactic elements only at PF (see Nunes, 2004 for a detailed discussion). It follows that grammatical transformations do not compound elementary operations (Chomsky, 1980). A syntactic element and its

copies form a nontrivial chain. The elements of a chain are distinguished in terms of their contexts (e.g., in [2] where the object *Seymour* is a sister of a verb while the subject *Seymour* is a sister of a TP (*was arrested Seymour in Shanghai*)).

In addition to the movement of NP to the Spec-TP position in passive constructions, IM also displaces the finite auxiliary verb in yes/no questions (4) and direct questions involving an interrogative pronoun (5), as well as wh-phrases in direct and indirect questions.

(4) Was Seymour arrested in Shanghai?

(5) In which city was Seymour arrested?

Example (5) involves three kinds of movement: (a) of an NP (*Seymour*), (b) of a head (*was*), and (c) of a wh-phrase (*in which city*). Each is performed by the same operation IM.

Given that IM is restricted to adjoining a phrase to the root, derivations cannot segregate applications of IM from those of EM, as was done in earlier versions of transformational grammar where movement transformations applied after the construction of an initial phrase marker. Thus in (3), the movement of the NP *Seymour* via IM occurs before the insertion of the complementizer *that* (the head of the Complementizer Phrase [CP]), as well as the verb *believe* and the pronoun *they*, via EM.

(6) They believe [_{CP} that [_{TP} Seymour_i was arrested Seymour_i in Shanghai]]

Because of this, derivations under bare phrase structure theory cannot define a canonical level of D-structure as in previous theories. Therefore, there is no level of D-structure under this theory.

Restricting the application of Merge to the root yields a version of 'strict cyclicity' (Chomsky, 1973; Freidin, 1978, 1999; H. Lasnik, personal communication, 2004), where a phrase that has become a constituent of another phrase cannot be altered by further applications of Merge (either EM or IM).

Constraints on Derivations and Representations

The operation Merge sketched in the previous section will by itself generate deviant constructions and therefore must be supplemented by further constraints to distinguish deviant from nondeviant utterances in a language. As has been demonstrated in research on dozens of languages over the past several decades, these constraints apply generally across languages. Constraints have been formulated in one of two ways, as conditions on the application of rules (i.e., on derivations) or as conditions on the representations constructed by rules. In current work, both kinds of

constraints have been formulated in terms of a notion of computational economy whereby the operation of rules and the representations they generate are optimally efficient in computing the relation between sound and meaning in linguistic expressions. In essence, economy conditions prohibit superfluous steps in derivations and superfluous elements in representations.

Full Interpretation

The fundamental condition on representations, Full Interpretation (FI) (Chomsky, 1986), requires that all elements in the representations at the interface levels, PF and LF, must be fully interpretable. Thus at LF, representations must contain no phonetic material under the assumption that such material cannot be interpreted by the cognitive modules that interface with LF. Similarly, PF representations must not contain any semantic information or syntactic constructs such as nontrivial chains, which cannot be interpreted at the sensorimotor interface.

Given FI, a derivation from N to LF must contain a point where phonetic material is separated from the syntactic representation and transferred to the phonological component of the grammar that constructs PF representations. This point, called Spell-Out, roughly corresponds to S-structure in previous analyses if there is only one point in a derivation. However, if there is multiple Spell-Out (Uriagereka, 1999; Chomsky, 2001), then classical S-structure cannot be constructed.

Applied at LF, FI eliminates constructions with superfluous arguments. For example, Merge could construct the following deviant example (7) in addition to the nondeviant examples (8a, 8b).

(7) *John gave Bill a book to Mary.

(8a) John gave Bill a book.

(8b) John gave a book to Mary.

The verb *gave* will assign argument functions (θ -roles) to the three NPs in both (8a) and (8b), but only three of the four NPs in (7). It can assign the indirect object θ -role to either *Bill* or *Mary*, but not to both. Therefore one of these NPs will not have a θ -role at LF, in violation of FI.

Case phenomena in languages are also governed by FI under the assumption that Case features are superfluous (hence uninterpretable) at both LF and PF and therefore have to be eliminated in the course of the derivation from N to LF. This can be illustrated with the following paradigm.

(9a) It is likely that he will graduate in June.

(9b) He is likely to graduate in June.

(9c) *It is likely he/him to graduate in June.

In (9a), the nominative Case pronoun *he* is the subject of the verb *graduate* and the matrix subject is the pleonastic element nonreferential *it*, which has no θ -role. The predicate adjective *likely* is a one-place predicate that assigns its single θ -role to the clausal complement *that he will graduate in June*. In (9b) the clausal complement of *likely* is infinitival rather than finite and the NP that is interpreted as the subject of the infinitival clause (*he*) has moved via IM to the subject position of the main clause. This movement is from a position that is assigned a θ -role to one that is not. This movement is obligatory because when it does not occur, as in (9c), the result is deviant. The difference between (9a, 9b) and (9c) is that the pronouns in the former occur in a position where they are licensed, whereas neither pronoun in the latter is licensed in that position. Once a Case feature is licensed it can be eliminated from the derivation and therefore will not show up at LF or PF. If it is not licensed, then it shows up at LF and PF in violation of FI. This analysis generalizes to overt NPs that do not show overt morphological Case (e.g., substitute the name *Adam* for the pronouns in [9] and the same result obtains), and therefore concerns a notion of abstract structural Case, which plausibly plays no role at PF (in contrast to overt morphological Case).

This analysis generalizes to the analysis of simple passives as discussed above. Thus compare (2) with the deviant (10).

(10) *It was arrested Seymour in Shanghai.

The pleonastic nonreferential *it* occurs in a non- θ -position and is licensed for Case as the subject of a finite clause. In contrast, *Seymour* occurs in a θ -position. Assuming that the reason the NP *Seymour* moves in (2) is to get to a Case-licensed position and thereby eliminate its Case feature, then the object position of a passive predicate must not be licensed for Case, in which case the NP *Seymour* violates FI at LF and PF because it still has whatever abstract Case feature it brings from the lexicon.

The abstract structural character of Case licensing can be more sharply illustrated in languages with more complex Case phenomena than English. Consider the following paradigm from Icelandic.

- | | | | |
|-------|------------|-----------------|-----------------|
| (11a) | Jóni | batnaði | veikin |
| | John-DAT | recovered-from | the-disease-NOM |
| (11b) | Ég tel [TP | Jóni hafa | batnað veikin] |
| | I believe | John- to | recovered- the- |
| | | DAT have | from disease- |
| | | | NOM |
| (11c) | [CP að [TP | PRO | batna |
| | | PRO-DAT | to-recover-from |
| | | veikin]] | er venjulegt |
| | | the-disease-NOM | is usual |

- (11d) *<sub>[CP að [TP Jóni batna
John-DAT to-recover-from
veikin]] er mikilvægt
the-disease-NOM is important</sub>

Although the verb *batna* ('recover from') selects a dative subject as a lexical property, (11c) illustrates that the null subject PRO can satisfy this requirement even though it has no phonetic content. Yet as (11d) shows, lexical satisfaction of the Case requirement is not sufficient since the lexically Case-marked NP cannot have phonetic content. The subject of the infinitive in (11d) is not structurally licensed for Case, in contrast to (11b) where the dative NP occurs in an 'exceptional Case-marking' context governed by the matrix verb *tel*. In (11b), the element that determines the actual morphological Case is distinct from the element that licenses abstract Case. Such examples demonstrate the need to distinguish morphological Case assignment from abstract structural Case licensing (see Freidin and Sprouse, 1991). Both requirements must be satisfied in order to remove uninterpretable Case features from the derivation.

The FI account of obligatory movement can be extended to wh-movement (i.e., movement to Spec-CP position) under the assumption that the head C contains a formal syntactic feature, which is uninterpretable and therefore must be eliminated in the course of the derivation. In (12), for example, the wh-phrase moves from a Case-marked object position to Spec-CP, which is Case-less.

- (12a) Which books did you borrow?
(12b) [_{CP} [_{NP} Which books]_i did_j [_{TP} you did_j borrow
[_{NP} which books]_i]]

It is also a movement from a θ -position to a non- θ -position. Since the movement must be overt (i.e., show up in PF), the formal feature involved must be uninterpretable at both LF and PF and therefore must be eliminated before Spell-Out.

In some languages (e.g., Chinese), wh-phrases do not move in overt syntax. Instead they move after Spell-Out i.e. covertly in LF (see Huang, 1981/1982). The covert movement of wh-phrases is also motivated by FI under the assumption that a wh-phrase constitutes a quasi-quantifier and as such must bind a variable created by movement at LF. Otherwise the quantifier without a variable to bind, a case of vacuous quantification, is uninterpretable. Presumably, Chinese lacks the formal feature that forces overt wh-movement.

Agreement features on verbs are also formal syntactic features that are considered uninterpretable at PF and LF and therefore can trigger a violation of FI unless they are eliminated before Spell-Out. If the agreement features on finite T enter the derivation

unvalued and are then valued by the agreement features of the head of Spec-TP, then unless an NP is merged in that position, the agreement features of T remain unvalued and violate FI at the interfaces. Thus it follows from FI that a finite clause must have a syntactic subject, even in languages where the subject is absent at PF. Thus agreement features on verbs can also be considered as a trigger for NP-movement.

Whether this analysis can be extended to cases of improper movement as in (13) is not as clear.

- (13a) *He is likely will graduate in June.
(13b) *_{[TP he_i is likely [_{TP} he_i will graduate in June]]}

When the pronoun *he* is first merged in the complement clause, it values the agreement features of the complement T. It is generally assumed that once a NP values the agreement features of one T, it cannot value the agreement features of another. Assume this to hold whether or not the NP is frozen in place after agreement applies. Then the agreement features on matrix T in (13) will remain unvalued, in violation of FI.

Not all improper movement cases involve a violation of FI. In (14), movement is from the subject of a finite clause to the subject of an infinitival clause.

- (14a) *It is easy he to believe has written a book.
(14b) *it is easy [_{TP} he_i to believe [_{TP} he_i has written
a book]]

The pronoun *he* moves from a Case-marked to a Case-less position. The Case features of the pronoun and the agreement features of the finite T are all eliminated from the derivation, so the result does not violate FI. However, notice that the uninterpretable Case and agreement features were eliminated prior to the movement of the pronoun.

One way to account for the deviance of (14) is to prohibit movements between two θ -positions by restricting the number of θ -roles a chain may have to one (see Freidin, 1978; Chomsky, 1981). An alternative is to prohibit the movement itself on the grounds that it is unmotivated by any grammatical factor, in contrast to legitimate phrasal movements. Since the pronoun *he* is both licensed for Case and assigned a θ -role in its initial position as subject of the complement of *believe*, there is no motivation for it to move further. From this perspective, movement is viewed as a 'last resort' to save a derivation from violating FI. Any movement that fails to meet this criterion is superfluous and therefore ruled out on general grounds of computational efficiency.

Locality of Movement

In addition to constraints on when and where a phrase must move and when it may not move, must also specify 'how far' a movement can be. Although

IM has the capability of moving a constituent indefinitely far away from its EM position, actual movement phenomena in languages demonstrate that certain syntactic boundaries may not be crossed; in effect, IM may not move a constituent 'too far' away. Thus movement operations are bounded within a particular syntactic domain. Compare, for example, the long-distance *wh*-movement in (9a), where the *wh*-phrase is interpreted as the verbal object in a clause (analyzed here as a Tense phrase (TP) contained in a Complementizer phrase [CP]) that is itself embedded in three other clauses, to the much shorter but impossible movements in (15b) and (15c).

- (15a) $[_{CP} \text{ who}_i \text{ did } [_{TP} \text{ you say } [_{CP} \text{ that } [_{TP} \text{ Mary knows } [_{CP} \text{ that } [_{TP} \text{ Bill told Sam } [_{CP} \text{ that } [_{TP} \text{ Rachel admires who}_i]]]]]]]]]$
- (15b) $*[_{CP} \text{ who(m)}_i \text{ did } [_{TP} [_{NP} \text{ your interest in who(m)}_i] \text{ surprise Mary}]]]$
- (15c) $*[_{CP} \text{ what}_j \text{ did } [_{TP} \text{ John wonder } [_{CP} [_{PP} \text{ to whom}]_i \text{ } [_{TP} \text{ Bill gave what}_j \text{ } [_{PP} \text{ to whom}]_i]]]]]$

In contrast to (15a), the movement in (15b) is intra-clausal rather than interclausal and the movement in (15c) only crosses one clause boundary instead of many.

The boundedness of IM can be accounted for in terms of a general principle of grammar that designates certain categories as bounding categories and sets a limit for bounding categories that may be crossed in a single-movement operation. This limit is fixed in UG at one bounding category. Two syntactic positions are said to be subjacent to one another if they are separated by no more than one bounding category. The principle for bounding, the Subjacency Condition, has the effect of prohibiting a single movement that crosses more than one bounding category.

(16) Subjacency Condition

Movement can only be to a subjacent position.

This principle can be formulated as a condition on chain links, in which case it constitutes a condition on representations rather than a condition on the application of rules (see Freidin, 1978; Browning, 1991). Alternatively, it has been argued that the Subjacency Condition must be formulated as a condition on the application of the movement rule (see Lasnik and Saito, 1984, 1992). Assuming the bounding categories for Subjacency are NP and TP, (16) rules out (15b) and (15c). To account for (15a), it is assumed that *wh*-movement operates locally when it can, so that *who* moves successively through each of the intervening Spec-CP positions, thereby avoiding a Subjacency violation. The condition applies more generally to

many other constructions involving movements across two NPs or across two TPs (see Freidin, 1992 for discussion and references). As a condition on representations, it applies at S-structure (see Freidin and Lasnik, 1981 for discussion) and therefore is incompatible with the multiple Spell-Out analysis mentioned above.

Bounding effects appear to vary across grammars. In some idiolects of Italian, it is possible to extract a relative pronoun out of an indirect question (see Rizzi, 1982).

- (17) $\text{il mio primo libro, } [_{CP} \text{ che}_i \text{ } [_{TP} \text{ credo } [_{CP} \text{ che}_i \text{ che } [_{TP} \text{ tu sappia } [_{CP} [_{PP} \text{ a chi}]_j \text{ } [_{TP} \text{ ho dedicato che}_i \text{ } [_{PP} \text{ a chi}]_j]]]]]]]$ *me è sempre stato molto caro*
'my first book, which I believe that you know to whom I dedicated, has always been very dear to me'

The sentence is well formed even though the movement between two trace positions crosses two TPs. This suggests that TP is not a bounding category in this idiolect of Italian. Nonetheless, a bounding effect occurs in Italian when a single movement of *che* crosses two CPs, as in (18).

- (18) $*\text{il mio primo libro, } [_{CP} \text{ che}_i \text{ } [_{TP} \text{ so } [_{CP} [_{PP} \text{ a chi}]_j \text{ } [_{TP} \text{ credi } [_{CP} \text{ che}_i \text{ che } [_{TP} \text{ abbia dedicato che}_i \text{ } [_{PP} \text{ a chi}]_j]]]]]]]$ *me è sempre stato molto caro*
'my first book, which I know to whom you believe that I dedicated, has always been very dear to me'

If CP is taken as a bounding category for this idiolect of Italian, then the second and final movement of the relative pronoun *che* violates Subjacency. Similar phenomena pointing to the same conclusion can be found in French and Spanish (see Sportiche, 1981 and Torrego, 1984). Thus the choice of TP vs. CP as a bounding category for Subjacency constitutes one parameter of UG. If both TP and CP are chosen as bounding categories, then Subjacency would prohibit interclausal *wh*-movement. Some languages (e.g., Russian) prohibit this kind of *wh*-movement, a fact that would follow from bounding theory if both TP and CP were bounding categories for Subjacency in these languages.

The Subjacency analysis remains the most general and perhaps the clearest account of syntactic island phenomena (i.e., construction from which the movement of constituents is prohibited). However, it may well be that the Subjacency Condition constitutes a false generalization, as have several other important proposals in the history of generative grammar. Moreover, the stipulation of bounding categories is essentially descriptive rather than explanatory and the principle itself is not easily interpreted in term of economy.

More recently, other analyses have been proposed that cover some of the empirical range of the Subjacency Condition. Chomsky (1993) proposes an economy principle that derivations involve the shortest steps (the Minimal Link Condition; see also Chomsky, 1995b), which prohibits (15c) on the grounds that the movement of *what* to the matrix Spec-CP is not the shortest movement possible because the movement from Spec-CP to Spec-CP is shorter. Chomsky (2001) proposes that subparts of derivations (phases) become inaccessible to further computation because they are transferred to PF and LF when complete. Details aside, under this proposal, *what* in (15c) would be inaccessible for movement because it is not at the edge of the CP phase. The proposals are problematic given that the movements they prohibit for English appear to be possible in several Romance languages (e.g., as in [17] above). Whether there is an alternative to the Subjacency analysis of locality with the same empirical coverage remains to be determined.

Binding Theory

Movement is not the only process that creates relations between pairs of positions within a sentence. Anaphoric expressions also create such relations via the interpretive link to their antecedent. Languages contain two types of anaphoric expressions: those that require the presence of an antecedent in the same sentence, called bound anaphors (i.e., reflexive pronouns and reciprocal expressions, henceforth anaphors), and pronouns (including anaphoric epithets like *the bastard*), which can occur in a sentence without an antecedent. Binding theory is concerned in part with the conditions under which anaphors and pronouns can be linked to an antecedent. (For more detailed discussion see Freidin, 1986; Lasnik, 1989; Koster and Reuland, 1992; and Chomsky and Lasnik, 1993.)

Bound Anaphors Formally, an antecedent binds an anaphoric expression when the anaphor is a constituent of the phrase that is merged with the antecedent (i.e., its sister). The paradigm for anaphors in (19) illustrates that all possible binding relations are not licit. (Again coindexing is used here and in what follows as a notational convenience.)

- (19a) simple sentence
 we_i help each other_i
 (19b) finite clause complement:
 (19b.i) $*we_i$ expect [_{CP} (that) [_{TP} each other_i
 will win]]
 (19b.ii) $*we_i$ expect [_{CP} (that) [_{TP} Mary will
 help each other_i]]

- (19c) infinitival complement:
 (19c.i) we_i expect [_{TP} each other_i to help Mary]
 (19c.ii) $*we_i$ expect [_{TP} Mary to help
 each other_i]

While all the anaphors in (19) are bound, only those in (19a) and (19c.i) are properly bound. Proper binding of anaphors falls under a principle of grammar, referred to as Principle A of the binding theory, which is given schematically in (20).

- (20) Principle A: an anaphor must be antecedent-bound in a local domain δ .

(20) constitutes a principle of UG which, by specifying the local domain δ , is instantiated in the grammar of a particular language as binding condition A.

The local domain for anaphors in English is the domain of an accessible SUBJECT, where SUBJECT stands for a syntactic subject or the agreement element associated with the head of finite TP (given that specifier-head agreement between a syntactic subject and the head T is mandatory in finite clauses). Any syntactic subject is accessible to an anaphor if it binds the anaphor. Thus (19a) and (19c.i) satisfy Condition A because the only accessible subject in each example is the main clause subject to which the anaphor is bound. In (19b.ii) and (19c.ii), the complement subject is accessible to the anaphor, but the anaphor is antecedent-free (i.e., not bound by an antecedent) in the domain of this 'subject'. Therefore, these examples violate Condition A. In (19b.i), the agreement element of the sentential complement constitutes an accessible SUBJECT and in that domain the anaphor is antecedent-free in violation of Condition A. (The precise formulation of the local domain δ for English is somewhat more intricate; see Chomsky, 1986; Freidin, 1986, 1992 for details.)

Cross-linguistically, we find some variation in the local domain δ for Condition A. In Korean, for example, the binding domain for reciprocals is that of any c-commanding NP. Thus within a single clause, a direct object may create a binding domain for the object of a PP in VP, as illustrated in (21) (from Hong, 1985).

- (21) kitil-i [_{NP} John-kwa Mary]-lil
 they-NOM John and Mary-ACC
 sero-eke sogehetta
 each other-to introduced
 they_i introduced [_{NP} John and Mary]_j to each
 other_{*i,j}

In the corresponding English example, either the subject or the direct object may be the antecedent of the reciprocal. Hence the binding domain for reciprocals in Korean is more restricted than that of English. This indicates that the binding domain for Principle

A is subject to parameterization. (See Freidin, 1992: Chap. 7 for a more detailed discussion of parameterized binding.)

Pronouns In contrast to bound anaphors, which must be antecedent-bound within a certain domain, pronouns cannot be bound within certain domains. The paradigm for pronoun binding corresponding to (19), given in (22), illustrates a complementarity between bound anaphors and bound pronouns where a bound anaphor can occur, a bound pronoun may not.

- (22a) simple sentence
 *the politician_i voted for him_i
 (22b) finite clause complement:
 (22b.i) Clara_i expects [_{CP} (that) [_{TP} she_i will win]]
 (22b.ii) Clara_i expects [_{CP} (that) [_{TP} Sam will help her_i]]
 (22c) infinitival complement:
 (22c.i) *Clara_i expects [_{TP} her_i to help Sam]
 (22c.ii) Clara_i expects [_{TP} Sam to help her_i]

Thus pronoun binding is accounted for by a principle of grammar complementary to Principle A, referred to as Principle B of the binding theory.

- (23) Principle B: a pronoun must be antecedent-free in local domain δ .

As with Principle A, Principle B is instantiated in the grammar of a particular language as binding condition B by specifying the local domain δ . As with anaphor binding, this domain varies across languages. For example, pronoun binding in the Icelandic sentence corresponding to (22c.ii) is not possible.

Given the complementarity of anaphors and bound pronouns illustrated in (19) and (22), it would appear that the specification of the local domain for pronoun binding is identical to that of anaphor binding. Although this complementarity holds for binding within TP and across TP boundaries, it collapses with respect to binding across NP boundaries, as examples (24–25) illustrate.

- (24a) they_i never discuss [_{NP} each other's_i work]
 (24b) they_i never discuss [_{NP} their_i work]
 (25a) every man_i heard [_{NP} a story about himself_i]
 (25b) every man_i heard [_{NP} a story about him_i]

These examples show that the domain statement for Condition B in English cannot be identical to that of Condition A, since if it were, then (24b) and (25b) should be unacceptable on a par with (22a) and (22c.i), contrary to fact.

These examples show that NP constitutes a binding domain for pronouns and the paradigm in (22) shows that TP constitutes another binding domain with one

exception – (22c.i) – where the matrix verb governs the complement subject across TP. One property that NP and TP share in common is that they both constitute domains in which predicates (e.g., verbs and derived nominals) may discharge their θ -roles (θ -domains, or complete functional constructs [in the terminology of Chomsky, 1986]). Example (22c.i) shows that the binding domain for pronouns is not the minimal θ -domain containing the pronoun, but rather the minimal θ -domain of the matrix verb, which is implicated in Case licensing of the pronoun. (For further discussion of the representation and interpretation of pronoun binding, see Lasnik, 1989.)

Universally, a pronoun cannot be anaphoric on an antecedent when it is in a structural binding relation to the antecedent (i.e., it is merged with a phrase that contains the antecedent), as in (26).

- (26) *she_i expects [_{CP} (that) [_{TP} Sam will help Clara_i]]

Example (26) can never be interpreted as a sentence about two people. Note that this is not a matter of the pronoun preceding its antecedent since (27) has the same linear property but not the same interpretive restriction.

- (27) His professor praised John in class.

Example (27) can be interpreted as a sentence about two people as well as a sentence about three. The constraint that accounts for the impossibility of the interpretation represented in (26) is designated in the literature as Principle C.

Conclusion

The material covered in this article is merely a sketch of some of the central topics being investigated within the principles and parameters framework. Ongoing research extends the empirical coverage of various subtheories, including parametric variation across languages. Current work also explores a number of promising alternative analyses for various parts of the theory. As in any area of rational inquiry, no assumption is immune to critical inquiry.

For further discussion of these and other topics within this approach to generative grammar, see van Riemsdijk and Williams, 1986; Lasnik and Uriagereka, 1988, 2005; Lasnik, 1999, 2003; Freidin, 1992; Epstein and Hornstein, 1999; and Epstein and Seely, 2002 among many others. Many of the topics under investigation have a long and rich history in generative grammar of the past half century: see Newmeyer, 1986; Freidin, 1994b for an historical account.

The principles and parameters framework of generative grammar attempts to establish the common biological basis for all the world's languages.

The current work in comparative grammar illuminates the human language faculty as an intricate computational structure of the human mind that is unique to the genetic endowment of our species. In this way work in contemporary comparative grammar continues to make a major contribution to the emerging cognitive science enterprise.

See also: Transformational Grammar: Evolution.

Bibliography

- Anderson S & Kiparsky P (eds.) (1973). *A festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.
- Baker M (1995). *The polysynthesis parameter*. New York: Oxford University Press.
- Baker M (2001). *The atoms of language*. New York: Basic Books.
- Browning M A (1991). 'Bounding conditions on representation.' *Linguistic Inquiry* 22, 541–562.
- Campos H & Kempchinsky P (eds.) (1995). *Evolution and revolution in linguistic theory*. Washington, DC: Georgetown University Press.
- Chomsky N (1964). *Current issues in linguistic theory*. The Hague: Mouton.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky N (1973). 'Conditions on transformations.' In Anderson & Kiparsky (eds.), 1973.
- Chomsky N (1976). 'Conditions on rules of grammar.' *Linguistic Analysis* 2, 303–351.
- Chomsky N (1980). 'On binding.' *Linguistic Inquiry* 11, 1–46.
- Chomsky N (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky N (1986). *Knowledge of language: its nature, origin, and use*. New York: Praeger.
- Chomsky N (1995a). 'Bare phrase structure.' In Campos H & Kempchinsky P (eds.), 1995.
- Chomsky N (1995b). *The minimalist program*. Cambridge: MIT Press.
- Chomsky N (2001). 'Derivation by phase.' In Kenstowicz (ed.), 2001.
- Chomsky N & Lasnik H (1993). 'The theory of principles and parameters.' In Jacobs J, von Stechow A, Sternefeld & Vennemann (eds.), 1993 [reprinted in Chomsky, 1995b].
- Epstein S & Hornstein N (eds.) (1999). *Working minimalism*. Cambridge: MIT Press.
- Epstein S & Seely D (eds.) (2002). *Derivation and explanation in the minimalist program*. Oxford: Blackwell.
- Freidin R (1978). 'Cyclicity and the theory of grammar.' *Linguistic Inquiry* 9, 519–549.
- Freidin R (ed.) (1991). *Principles and parameters in comparative grammar*. Cambridge: MIT Press.
- Freidin R (1992). *Foundations of generative syntax*. Cambridge: MIT Press.
- Freidin R (1994a). 'Generative grammar: principles and parameters framework.' In Asher R E (ed.) *The encyclopedia of language and linguistics*, vol. 3. Oxford: Pergamon Press. 1370–1385.
- Freidin R (1994b). 'Conceptual shifts in the science of grammar: 1951–1992.' In Otero C P (ed.), 1994, vol 1.
- Freidin R (ed.) (1996). *Current issues in comparative grammar*. Dordrecht: Kluwer.
- Freidin R (1997). 'The minimalist program.' *Language* 73, 571–582.
- Freidin R (1999). 'Cyclicity and minimalism.' In Epstein & Hornstein (eds.), 1999.
- Freidin R & Sprouse R A (1991). 'Lexical case phenomena.' In Freidin (ed.), 1991.
- Hong S (1985). *A and A-binding in Korean and English*. Ph.D. diss., University of Connecticut.
- Huang C-T J (1981/1982). 'Move *wh* in a language without *wh*-movement.' *The Linguistic Review* 1, 369–416.
- Jackendoff R (1972). *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- Jackendoff R (1990). *Semantic structures*. Cambridge: MIT Press.
- Jacobs J, von Stechow A, Sternefeld W & Vennemann T (eds.) (1993). *Syntax: an international handbook of contemporary research*. Berlin: de Gruyter.
- Jaeggli O & Safir K (eds.) (1989). *The null subject parameter*. Dordrecht: Kluwer.
- Kayne R (1975). *French syntax: the transformational cycle*. Cambridge: MIT Press.
- Kenstowicz M (ed.) (2001). *Ken Hale: a life in language*. Cambridge: MIT Press.
- Koster J & Reuland E (eds.) (1992). *Long-distance anaphora*. Cambridge: Cambridge University Press.
- Lasnik H (1989). *Essays on anaphora*. Dordrecht: Kluwer.
- Lasnik H (1999). *Minimalist analysis*. Oxford: Blackwell.
- Lasnik H (2003). *Minimalist investigations in linguistic theory*. London: Routledge.
- Lasnik H & Saito M (1984). 'On the nature of proper government.' *Linguistic Inquiry* 15, 235–289.
- Lasnik H & Saito M (1992). *Move α* . Cambridge: MIT Press.
- Lasnik H & Uriagereka J (1988). *A course in GB syntax: lectures on binding and empty categories*. Cambridge: MIT Press.
- Lasnik H & Uriagereka J (2005). *A course in minimalist syntax: foundations and prospects*. Oxford: Blackwell.
- Newmeyer F (1986). *Linguistic theory in America* (2nd edn.). Orlando, FL: Academic Press.
- Nunes J (2004). *Linearization of chains and sideward movement*. Cambridge: MIT Press.
- Otero C P (ed.) (1994). *Noam Chomsky: critical assessments*. London: Routledge.
- Quicoli A C (1976a). 'Conditions on clitic-movement in Portuguese.' *Linguistic Analysis* 2, 199–223.
- Quicoli A C (1976b). 'Conditions on quantifier movement in French.' *Linguistic Inquiry* 7, 583–607.
- Rizzi L (1982). *Issues in Italian syntax*. Dordrecht: Foris.
- Rizzi L (2000). *Comparative syntax and language acquisition*. London: Routledge.

- Ross J R (1967). *Constraints on variables in syntax*. Ph.D. diss., MIT. Published as 1984 *Infinite syntax!*, Ablex, Norwood, NJ.
- Rouveret A & Vergnaud J-R (1980). 'Specifying reference to the subject: French causatives and conditions on representations.' *Linguistic Inquiry* 11, 97–202.

- Sportiche D (1981). 'On bounding nodes in French.' *The Linguistic Review* 1, 219–246.
- Torrego E (1984). 'On inversion in Spanish and some of its effects.' *Linguistic Inquiry* 15, 103–129.
- Uriagereka J (1999). 'Multiple spell-out.' In Epstein & Hornstein (eds.), 1999.

Private Language Argument

D Byrne, University of Birmingham, Birmingham, UK

© 2006 Elsevier Ltd. All rights reserved.

The private language argument is the strand of Ludwig Wittgenstein's later work (in particular, 1953: §§243–315) in which he endeavored to demonstrate the impossibility of a language that only one person could understand – or, at any rate, as Wright (1986) suggested, that only one person could have *reason to believe* he or she understood. Early commentators, although divided over the argument's cogency, regarded it as perhaps the most powerfully subversive weapon in Wittgenstein's armory. Fashions have changed, however, and nowadays few philosophers seem to agree even on so much as whether the argument exists and, if it does, how it works. Critics satirize the legendary argument as a kind of perpetually elusive 'Philosopher's Stone' of woolly-minded pragmatism, and even Wittgensteinians tend nowadays to downplay the significance of any single *argument*, preferring to emphasize themes expounded by Wittgenstein in the surrounding passages (see Hacker, 1986; McDowell, 1989; McGinn, 1997).

The Argument's Target

Wittgenstein's interest is probably not linguistic *per se*: his intention is to attack a general conception of mind, usually labeled Cartesian. According to this picture, introspection affords thinking subjects an especially direct, observational access to their own mental states, whereas their access to the mental states of others is indirect and theoretical – a matter of inference from their behavior. This conception seems particularly compelling in regard to sensations and other phenomenal experiences.

Wittgenstein invites us to consider a subject – let us call him *PD* (Private Diarist) – who undertakes to name various types of sensations and record their occurrences in a diary. Cartesianism would seem to suggest that, to inaugurate names of sensation-types in this way, it would suffice for *PD* simply to "speak, or write the sign down, and at the same time ...

concentrate ... attention on the sensation – and so, as it were, point to it inwardly" (Wittgenstein, 1953: §258). The Cartesian conception also seems to suggest that subjects other than *PD* could have no reason to believe they understood these terms, so that a language containing them would be *private* to *PD* in the sense intended. By arguing that such a language is impossible, then, Wittgenstein attacks the Cartesian conception of mind that underpins it.

The Cartesian conception continues to dominate our intuitive, 'pretheoretical' view – at least of those mental states that enjoy a distinctive phenomenology. This alone ensures that it really matters whether there is a cogent private language argument. However, before addressing that question, it is also worth briefly considering which contemporary theories – of philosophy, psychology, linguistics, and so on – are threatened by the argument. Wittgensteinians frequently bemoan the widespread infection of mainstream theory in these disciplines by Cartesianism, but although there are reasonable general grounds for this complaint, it is actually quite difficult to find theories that are threatened by the private language argument *per se*.

The 'representational theory of mind' that underwrites much recent work in cognitive science and linguistics might seem the most obvious candidate. Indeed, Fodor (1975) has provocatively placed himself on the firing-line on its behalf. According to the representational theory, cognition involves the manipulation of symbols that, like linguistic expressions, enjoy representational contents, yet that are private in the sense that those contents are not determined by anything like public conventions or criteria. As is discussed below, there is an interpretation of the private language argument that would place the notion of contentful symbols that are private in this sense in its sights, but it is not the most plausible interpretation. Meanwhile, we can already see that there is no *prima facie* tension between the representational theory and the argument's conclusion. For one thing, Fodor did not suppose that representations' contents are conferred by anything like acts of introspective ostension. Even more to the point,

an advocate of the representational theory is entitled to insist that there could be good (empirical) reasons to suppose that two subjects enjoyed representations with the same content (and so to suppose that, in a sense, subjects ‘understood’ the same representations).

Much more vulnerable to challenge are the ‘direct reference’ theories of ‘phenomenal concepts’ recently defended by Loar (1997), Chalmers (2003), and others. The key idea here is that subjects enjoy concepts of the phenomenal qualities of their own conscious experiences whose contents are exhausted by their references. Some of these theorists suggest that such concepts are inaugurated through acts of introspective demonstration, and given the distinctive semantics they are proposed to enjoy, it is difficult to see how such concepts could fail to be private in something like the sense at issue here.

Supplementary Strands

Most of Wittgenstein’s remarks in the passages generally considered to contain the private language argument do not directly address the issue identified above as central, but instead are concerned with the meanings of expressions used in public languages to talk about sensations and other mental states.

Wittgenstein held quite generally that understanding an expression is a matter of knowing and following rules that regulate how it is used in various language-involving activities – ‘language games’ – in which speakers participate. One thing Wittgenstein did in these passages is to emphasize the third-person uses of (public language) sensation terms. Understanding words like ‘toothache’ and ‘pain’ involves, *inter alia*, knowing what kinds of behavioral criteria sanction their third-person attribution. Therefore, at any rate, that understanding cannot be just a matter of demonstrative first-person acquaintance with pains, as a thoroughgoing Cartesian might assume. Of course, a Cartesian might reply by conceding the point about public language sensation terms, but insisting nonetheless on the possibility of distinct, privately inaugurated terms lacking third-person applications altogether.

Elsewhere in these passages, Wittgenstein made characteristic proposals to the effect that what seem superficially to be assertions may in fact be speech acts of other kinds: warnings, exclamations, ‘performatives,’ etc. One example is his suggestion that utterances involving (public language) sensation terms can be used to express sensations, rather than to describe them: “Here is one possibility: Words are connected with the primitive, the natural, expressions of the sensation and used in their place” (§1953: 244). Considering sentences that might seem to articulate

the most fundamental features of sensations, such as ‘Sensations are private,’ Wittgenstein suggested that these are used to express the rules that govern the terms involved, rather than to make substantial assertions: ‘Sensations are private’ is comparable to “One plays patience by oneself” (1953: §248). These points are interesting and important; however, even if Wittgenstein is right that apparently indicative sentences featuring public language sensation terms can have these nonassertoric uses, this does not refute the Cartesian’s basic belief in the possibility of a different language, in which he can privately make assertions about his own sensations.

Rules and Communities

Why, then, does Wittgenstein repudiate the possibility of a private language? His basic contention seems to be that a subject who undertook to inaugurate names for sensation-types he experiences by defining them through introspective demonstrations would not succeed in imbuing the names with genuine meaning – even ‘for himself.’ Yet, why not?

One point Wittgenstein makes is that, normally, inaugurating a new name requires a lot of background “stage-setting” (1953: §257) of a kind unavailable here. For example, parents’ ability to name their children depends on a background of conventions concerning the roles played by persons’ names in the various language-games in which they feature. However, the Cartesian is likely to respond that PD might instate ‘private conventions’ of suitable kinds. Whether this is possible is really just the question at issue, so Wittgenstein owes more of an argument than is provided by a catalogue of differences between the private and public cases.

Just such an argument seems to be offered in this famous passage:

I impress on myself the connexion between the sign and the sensation. – But ‘I impress it on myself’ can only mean: this process brings it about that I remember the connexion *right* in the future. But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can’t talk about ‘right’ (1953: §258).

Suppose that PD experiences a sensation and attempts demonstratively to label the type of sensation it represents as ‘S.’ If ‘S’ is to be genuinely meaningful, there must be a difference between PD’s later (1) correctly applying it to a sensation that really is of the same type as the first, and his (2) misapplying it by mistakenly applying it to a sensation that is really of a different type. Wittgenstein’s contention seems to be that no such distinction can be sustained in this case,

because PD has no ‘criterion of correctness’ to go on independent of his own inclinations. The question the Cartesian will want to press, however, is precisely why such a criterion must be in this way independent. The argument should not be based on an assumption that PD’s judgments about whether a certain sensation is of the same phenomenal kind as the one he previously labeled ‘S’ are *unreliable* – that is on a general skepticism about PD’s memory – for such an assumption would, once again, simply beg the question against the Cartesian.

As we saw above, Wittgenstein holds that linguistic understanding involves knowledge of rules regulating the use of terms. A striking feature of such knowledge seems to be its generality, because understanding a term involves the ability to use it in many different kinds of situations. He is particularly preoccupied with the question of how the contents of rules can be in this way general and, yet at the same time, apparently constitute objects of conscious awareness: things that speakers can “grasp in an flash” (1953: §138). Some commentators maintain that Wittgenstein proposes to resolve this tension by, in effect, denying that meanings really are objects of conscious awareness and by holding that a speaker’s linguistic understanding is in the end simply a matter of her using terms in the same ways as other members of her linguistic community. According to this view, the patterns of use upon which members of a speech community converge engender ‘criteria’ of correct use that are in large measure independent of the judgments of individual members. Thus, in this picture, linguistic expressions generally are governed by criteria of correctness independent of individual speakers’ impressions, and Wittgenstein’s injunction that ‘S’ would be meaningful only if – *per impossible* – it could be so governed is just a special case of this general doctrine. Notice furthermore that if representation in general rests on this kind of communal convergence, then the private language argument threatens the representational theory of mind, as well as the Cartesian position discussed here.

The trouble with this interpretation of the private language argument is that the general ‘communitarian’ conception of rule following and meaning on which it rests is implausible. Moreover, commentators generally reject its attribution to Wittgenstein. (Wright (1980) explored the communitarian interpretation in detail and offered some defense. Kripke’s (1982) notorious interpretation may be considered a version of the communitarian one, its distinctive spin being the meaning-skepticism that Kripke also attributed to Wittgenstein. See Baker and Hacker (1984) and McDowell (1984) for forceful arguments against both the view and its attribution to Wittgenstein).

Radical Evidence-Independence and the ‘Theory Response’

I have argued that, although the ‘supplementary strands’ of Wittgenstein’s discussion highlight numerous important contrasts between ordinary public language sensation terms and the names that PD undertakes to introduce, these considerations do not motivate Wittgenstein’s contention that PD *cannot* imbue the latter with genuine meaning. There is an influential conception of rule following and meaning that does provide such a motivation, but it is widely rejected, both as a theory of meaning and an interpretation of Wittgenstein. In this final section, I describe a different way to defend Wittgenstein’s insistence that PD cannot inaugurate a meaningful term in the manner considered.

The crucial issue is that the Cartesian conception seems to imply that PD’s beliefs about whether a current sensation is of the type he earlier labeled ‘S’ could be correct or incorrect, even if no evidence bearing on the matter were available. Commitment to this possibility implies an extremely strong form of realism or ‘evidence-independence’ both about the sensations at issue and also (this is often overlooked) the intention – the baptismal ‘undertaking’ – in terms of which PD is proposed to have inaugurated ‘S’.

Advocates of the argument are often accused of resting illicitly on a verificationist and/or behaviorist assumption, and it is true that verificationism (in respect of mental states) and behaviorism are at odds with this radically realist picture. However, it is not necessary to subscribe to either of those views to feel uncomfortable about the degree of evidence-independence to which the Cartesian realist seems committed here. The Cartesian holds not only that there may be truths about whether a sensation that PD has is of the type he earlier labeled ‘S’, for which there is no conclusive evidence, but also that such truths may obtain in the absence of any evidence whatsoever – even of the entirely nonbehavioral, introspective kind associated with the Cartesian notion of privileged introspective access. Wittgensteinians, such as Wright (1991), have argued forcibly that this degree of evidence-independence is more than our ordinary psychological notions – particularly those of intentions – can bear. Consider, for example, the idea that a subject *believes that goldfish are highly dangerous*, even though after careful reflection, she is convinced that she does not believe that, and even though all of her behavior – recorded over as long a temporal interval and in as much detail as you like – demonstrates no inclination whatsoever to avoid goldfish, to protect her loved ones from them, etc. If this idea seems

absurd, then it looks as though we should also reject the radical evidence-independence of baptismal 'undertakings' to which the Cartesian seems committed.

Actually though, there is a way to question whether the Cartesian needs to be quite as rampantly realist as this. I have been working with the assumption that the only 'criterion of correctness' PD can apply is his own introspectively grounded impression of whether a current sensation is like the one he baptized 'S.' Yet, suppose PD has reflected on his sensations enough to formulate (in his private language) theories about them, concerning, for example, which types of sensations tend to cooccur during given temporal intervals and which not. Blackburn (1984) and others have suggested that commitment to well-confirmed theoretical hypotheses of this kind could give PD reason to disregard a phenomenal appearance at odds with it, just as a scientist might disregard the appearance of an experimental result at odds with accepted theory. If this is right, then PD might reasonably judge that an earlier application of 'S' to a sensation had been in error, even though his impression at the time had been that it was just like the one he initially associated with the term. This 'fallibilist' Cartesian can hold that there can be evidence of (sincere) misapplication (or indeed, of (sincere) correct application) of terms in a private language and so avoid commitment to the degree of evidence-independence described above.

The 'argument from radical investigation-independence' is the most formidable version of the private language argument, and the 'theory response' represents the most promising rejoinder. However, the response has not been developed in the kind of detail a conclusive adjudication would require. Wright (1986) delved deeper than most, arguing that the response depends on the assumption that PD's 'theory' satisfies various formal constraints that theories codifying patterns of cooccurrence and non-cooccurrence of sensations of given types over given temporal intervals are very unlikely to satisfy. However, the Cartesian is likely to hope that even if Wright is correct about theories of this particular, rather rarefied sort, a more general integration of private phenomenal concepts into a subject's thinking might engender a structure that does satisfy those constraints. (For some recent critical reaction to Wright (1986), see Bain (2004) and Byrne (in preparation).) As things stand then, the issues surrounding Wittgenstein's famous argument remain almost as open as they were a half-century ago.

See also: Concepts; Intention and Semantics; Mentalese; Ordinary Language Philosophy; Realism and Antirealism; Rules and Rule-Following; Verificationism.

Bibliography

- Bain D (2004). 'Private languages and private diarists.' *Philosophical Quarterly* 54, 427–434.
- Baker G & Hacker P (1984). *Scepticism, rules and language*. Oxford: Blackwell.
- Blackburn S (1984). 'The individual strikes back.' *Synthese* 58, 281–301.
- Byrne D (in preparation). 'The contents of phenomenal concepts.'
- Chalmers D (2003). 'The content and epistemology of phenomenal belief.' In Smith Q & Jokic A (eds.) *Consciousness: new philosophical perspectives*. Oxford: Oxford University Press.
- Fodor J (1975). *The language of thought*. New York: Crowell.
- Hacker P (1986). *Insight and illusion: themes in the philosophy of Wittgenstein* (rev. edn.). Oxford: Oxford University Press.
- Hacker P (1990). *Wittgenstein: meaning and mind: an analytical commentary on the philosophical investigations* (vol. 3). Oxford: Blackwell.
- Jones O R (ed.) (1971). *The private language argument*. London: Macmillan.
- Kripke S (1982). *Wittgenstein on rules and private language*. Oxford: Blackwell.
- Loar B (1997). 'Phenomenal states (revised version).' In Block N, Flanagan O & Güzelde G (eds.) *The nature of consciousness: philosophical debates*. Cambridge, MA: MIT Press.
- McDowell J (1984). 'Wittgenstein on following a rule.' *Synthese* 58, 325–363. (Reprinted in McDowell J (1998). *Mind, language and reality*. Cambridge, MA: Harvard University Press.)
- McDowell J (1989). 'One strand in the private language argument.' *Grazer Philosophische Studien* 33/34, 285–303. (Reprinted in McDowell J (1998). *Mind, language and reality*. Cambridge, MA: Harvard University Press.)
- McGinn M (1997). *Wittgenstein and the philosophical investigations*. London: Routledge.
- Pagin P (2000). 'Sensation terms.' *Dialectica* 54(3), 176–199.
- Pears D (1988). *The false prison* (vol. II). Oxford: Oxford University Press.
- Wittgenstein L (1953). *Philosophical investigations*. Oxford: Blackwell.
- Wright C (1980). *Wittgenstein on the foundations of mathematics*. London: Duckworth.
- Wright C (1986). 'Does *Philosophical Investigations* §§ 258–260 suggest a cogent argument against private language?' In Pettit P & McDowell J (eds.) *Subject, thought and context*. Oxford: Oxford University Press. (Reprinted in Wright C (2001). *Rules to infinity*. Cambridge, MA: Harvard University Press.)
- Wright C (1991). 'Wittgenstein's later philosophy of mind: sensation, privacy and intention.' In Puhl K (ed.) *Meaning scepticism*. Berlin and New York: De Gruyter. (Reprinted in Wright C (2001). *Rails to infinity*. Cambridge, MA: Harvard University Press.)

Proper Names: Philosophical Aspects

M Reimer, University of Arizona, Tucson, AZ, USA

© 2006 Elsevier Ltd. All rights reserved.

What Are Proper Names?

Intuitively, we can think of proper names as linguistic ‘tags.’ In the terminology of John Stuart Mill (1843), although proper names denote, they do not connote; they do not ‘imply attributes.’ For instance, when a speaker uses the name ‘Kripke’ to refer to the famous philosopher, he or she does not thereby imply anything about Kripke; he or she simply refers to the man. This intuitive construal of proper names is helpful insofar as it provides a rough idea of what philosophers are theorizing about when they claim to be theorizing about ‘proper names.’ Nevertheless, Mill’s characterization is not uncontroversial for reasons discussed later.

Two Central Issues: Meaning and Reference

Proper names raise two central philosophical questions. First, what is the meaning of a proper name? What (in other words) is its contribution to the meanings of the sentences in which it occurs? ‘Vienna’ presumably contributes something to the meaning of ‘Vienna has great pastries.’ What exactly does it contribute? Second, how does a proper name refer? How (in other words) does it ‘attach to’ its bearer? When I say that ‘Rumsfeld is in the news again,’ the subject term of the sentence I utter ‘picks out’ a certain Donald Rumsfeld, thereby enabling me to talk about him. What makes this possible? Some philosophers (including Frege, 1892/1962) believe that an answer to the first question will yield an answer to the second. Let us therefore begin with the first question.

Theories of Meaning

Since the rise of analytic philosophy in the early decades of the 20th century, two theories of meaning have dominated the literature on proper names: Millian and description theories. Description theories are motivated largely by their ability to handle the very sorts of cases that prove problematic for Millian theories. Let us therefore begin with the latter.

Millian Theories

According to Millian theories, the meaning of a name is simply its bearer. In other words, a name contributes its bearer (and only its bearer) to the meaning of any sentence in which it occurs. Thus, Vienna itself is what

‘Vienna’ contributes to the meaning of ‘Vienna has great pastries.’ Such theories are called ‘Millian’ after John Stuart Mill, who (as noted previously) famously claimed that proper names denote but do not connote. Thus, insofar as such expressions can be said to ‘mean’ anything at all, they mean what they denote.

Contemporary Millians, and philosophers of language generally, often prefer to talk in terms of the ‘proposition expressed,’ where this is understood as meaning (roughly) what is said by the literal utterance of a sentence. Thus, ‘Vienna’ is said to contribute Vienna itself to the proposition expressed by any sentence in which that name occurs. Vienna is said to be the ‘semantic value’ of the name. In this way, Millians can avoid the odd-sounding claim that names ‘mean’ what they refer to. Names do not ‘imply attributes’; they do not ‘connote.’ In that sense of meaning, names do not mean anything at all. Their semantic function is simply to denote and their denotations are what they contribute to the proposition expressed. Recent advocates of Millianism, both of whom talk in this vein, include Nathan Salmon (1986) and Scott Soames (1987).

A central virtue of Millianism is its intuitive plausibility. Intuitively, ‘Dick Cheney’ refers to Dick Cheney without thereby characterizing him in any way. This taglike feature of proper names appears to be what distinguishes such expressions from much of the rest of meaningful language. Ask someone what ‘Tucson’ means and you might well get a puzzled look followed by (something like) “It doesn’t really ‘mean’ anything, it’s the name of a city in Arizona!” In contrast, you might well get a response to a question about the meaning of ‘city.’ It means (you might be told) something like ‘metropolitan area.’

The central objections to Millianism concern special cases that the theory seems unable to accommodate: cases involving identity sentences, propositional attitude attributions, and ‘empty’ names (names without bearers). These are the very same cases that provide the central motivation for description theories.

Description Theories

Description theories (associated with Frege, Russell, and, more recently, Searle) claim that the meaning of a proper name is the descriptive or (more generally) conceptual content with which it is associated. For Frege (1892/1962), this associated content (which he called ‘sense’) is the ‘mode of presentation’ of the object referred to by the name. It represents the speaker’s ‘take’ or perspective on the name’s referent. For others (e.g., Russell, 1912/1917) the associated content is said to be a matter of what the speaker ‘has in

mind' when using a name. Description theorists sometimes talk as though associated content can be captured via a single description, one shared by members of the linguistic community at large (Russell, 1905, 1919). Yet both Frege (1892/1962) and Russell (1912/1917) acknowledged that the associated content could vary from speaker to speaker. For Russell (1912/1917), and perhaps for Frege as well, such content could even vary for an individual speaker over time. According to contemporary descriptivist Searle (1958, 1983), to the extent that the sense of a proper name can be captured linguistically, it is captured by a 'cluster' of descriptions the individual speaker associates with the name.

The virtues of description theories have been much touted in the philosophical literature. Specifically, such accounts appear to solve a variety of puzzles that can be illustrated by the following sentences:

- (1) Hesperus is Phosphorus.
- (2) Alice believes that Samuel Langhorne Clemens has a funny middle name.
- (3) Sherlock Holmes lives on Baker Street.
- (4) Santa Claus does not exist.

Sentences such as (1) are known as 'identity sentences,' sentences such as (2) are known as 'propositional attitude attributions' (because they attribute an attitude, such as belief or doubt, to an agent). Sentences (3) and (4) contain 'empty' names, and the latter is known as a 'negative existential' because it denies the existence of something.

Sentences such as these are puzzling on the assumption that Millianism is true. Suppose the meaning of a proper name is simply its bearer. Then (1), if true, is trivially true, and (2) is synonymous with (5):

- (5) Alice believes that Mark Twain has a funny middle name.

Whereas (3) is meaningless (due to a subject term without a bearer), (4) is nearly contradictory: If the subject term is meaningful (if it has a bearer), the sentence is false. Intuitively, however, all of this is mistaken. (1) is both true and informative, and (2) and (5) are not synonymous because the former might be true while the latter is false. (3) is meaningful and (4) is not only meaningful but also true. How can this be? Suppose we construe the meaning of a proper name as the descriptive content with which it is associated. Then proper names are, in effect, descriptions. Consistent with this general picture, let us adopt the following stipulations: 'Hesperus' means 'the brightest star in the evening sky,' 'Phosphorus' means 'the brightest star in the morning sky,' 'Samuel Langhorne Clemens' means 'my eccentric neighbor

Sam,' 'Sherlock Holmes' means 'the famous London detective featured in the Conan Doyle stories,' and 'Santa Claus' means 'the jolly bearer of Christmas gifts who lives at the North Pole.' Given these assumptions, we can explain the informativeness of (1), the fact that (2) and (5) mean different things, the meaningfulness of (3), and the truth of (4). To see this, one need only substitute for the proper name the description that captures the name's associated content (e.g., 'The brightest star in the evening sky is the brightest star in the morning sky'). The resultant sentences arguably have the same meaning properties as the original sentences – a fact that lends credence to descriptive accounts of proper names. (For details, see Frege (1892/1962) and Russell (1905), the latter of which contains Russell's analysis of the sorts of sentences that result when a proper name is replaced with a definite description.)

Contemporary Millians (including Salmon, 1986) are not impressed by the description theorist's analysis of sentences such as (1)–(5) and have responded that their own theories can easily account for the intuitive data. The basic idea is to claim that the intuitions surrounding such sentences are the result of mistaking what a speaker might communicate by their utterance with what the sentences themselves literally express (which is just as Millianism claims). Moreover, description theories of meaning are not problem-free. For instance, if a name 'N' means what some description 'the *F*' means, then why doesn't '*N* is the *F*' (e.g., 'Cheney is the current U.S. vice president') sound as trivial (in 2004) as 'bachelors are unmarried'? (For this and other problems, see Devitt and Sterelny, 1999.)

Theories of Reference

Some philosophers have thought that description theories of meaning can be extended to account for reference. These philosophers advocate description theories of reference, the *locus classicus* of which is Frege's (1892/1962) 'On sense and reference.' Such theories are no longer as popular as they once were. Since the 1970's, causal theories of reference have provided an attractive alternative to description theories. The causal theory (as originally sketched by Kripke in a series of lectures later published as *Naming and necessity* (Kripke, 1980)) arose in large part out of dissatisfaction with description theories. Let us therefore begin with description theories of reference.

Description Theories

According to any such theory, the referent of a proper name is determined by the descriptive content with

which it is associated; in other words, the referent is whatever 'satisfies' the associated content. According to some philosophers (Russell, 1905, 1919), the descriptive content can generally be captured linguistically; others (Searle, 1983) disagree. (Here, differences between versions of description theories of meaning are mirrored.)

The motivation for description theories is both intuitive and methodological. Ask a speaker to what he or she is referring when he or she uses a particular name and the speaker will likely provide a description in response. Ask me to what I am referring when I talk of 'Mount Everest' and I might well respond with an appropriate description, such as 'the tallest mountain in the world.' Such intuitive considerations at least suggest that a proper name (as used by a speaker) refers to its bearer in virtue of that entity's satisfying some descriptive criterion associated with the name. The theory also has a clear methodological virtue when coupled with a description theory of meaning. In this way, a single phenomenon (associated conceptual content) explains both meaning and reference.

The central objections to the theory include the following two, both from Kripke (1980). First, a name can refer uniquely to its bearer even if the associated descriptive content is satisfied by more than one object/individual. 'Cicero' refers uniquely to the famous Roman orator also known as 'Tully' even if the associated descriptive content is simply 'a Roman orator.' This suggests that something more than (or other than) associated descriptive content is needed to determine reference. Second, a name can refer to its bearer even if the associated descriptive content is uniquely satisfied by some other object/individual. 'Einstein' refers to the famous German physicist even if the associated descriptive content (e.g., 'the inventor of the atomic bomb') describes someone else (Oppenheimer). This suggests that something other than descriptive content determines reference. (For further objections to the description theory, see Devitt and Sterelny, 1999.)

Many philosophers (but not all, including Searle, 1983) have viewed Kripke's objections to description theories of reference as decisive. As a result, other theories have been developed, including, most notably, causal theories of reference, the first of which was sketched in Kripke's (1980) *Naming and necessity*.

Causal Theories

A causal theory of reference claims that the link between a name and its bearer is (at least generally) causal in nature. Following Devitt and Sterelny (1999), one can think of such theories as (ideally) having two components: a theory of reference 'fixing'

and a theory of reference 'borrowing.' The former explains how a name first gets attached to a certain entity; the latter explains how members of the linguistic community at large are subsequently able to use the name to refer to that entity. So construed, causal theories are easily illustrated. My daughter has a Chihuahua named 'Cocoa.' When she first brought him home, she decided to call him 'Cocoa' because of his chocolate color. She proceeded to call him 'Cocoa,' and Cocoa he thereby became. Reference was accordingly 'fixed' and all those present at the initial applications of the name thereby acquired the ability to use the name. The ability was acquired in virtue of having perceived (and thus been causally affected by) these initial applications. My daughter subsequently told her friends about her Chihuahua (using his name to do so) and thereby enabled those same friends to refer to him via 'Cocoa.' In this way, the causal (perceptual) chain of communication, originating in the initial applications of the name, is extended. My daughter's friends can extend the chain further, enabling still others to refer to Cocoa via 'Cocoa,' simply by using his name in their (attentive) presence. Reference is thereby 'borrowed.'

Although reference is generally causal on this picture, exceptions are acknowledged. It is uncontroversial that there are proper names whose reference is fixed by description. As Kripke (1980) noted, when 'Neptune' was introduced into the language by the astronomer Leverrier, its reference was stipulated to be 'the planetary cause of Uranus's orbital perturbations.' (See Evans, 1979, for a discussion of such 'descriptive' names.)

A central virtue of causal theories is that they avoid the problems of description theories. They claim, in effect, that associated descriptive content is irrelevant to reference. A second (no less important) virtue is that such theories provide a simple and intuitive answer to the question, 'How does language attach to reality?' In a word: causality. (See Devitt and Sterelny, 1999, for details.)

Nevertheless, causal theories are not without their problems. Gareth Evans, in his classic paper 'The causal theory of names' (1973), presented a telling objection to the theory, as originally sketched by Kripke. According to Evans, the causal theory fails to account for the fact of reference change. As Evans pointed out, 'Madagascar' once referred to a portion of the African mainland, but now it refers to the island. Evans argued that the causal theory cannot explain how such changes could occur. (But see Devitt and Sterelny, 1999, for a compelling rejoinder.) In developing an alternative that accounts for reference change, Evans was led to a hybrid theory, one that combines elements from both descriptive and causal theories.

Hybrid Theories

According to Evans's hybrid theory, a name refers to the dominant causal origin of the descriptive information with which it is associated. The intuitive pull of the theory is easily illustrated. Suppose a tattered and centuries-old manuscript is discovered. On its cover is emblazoned the name 'Melford Bibbins.' The manuscript is widely discussed among academics, and its author, known to all as 'Melford Bibbins,' becomes famous for his controversial philosophical views. Intuitively, the individual being referred to is the author of the manuscript, even if it turns out that 'Melford Bibbins' was the name of the scribe to whom the author dictated his work. Evans's theory gets it right: the author of the manuscript is the causal origin of the descriptive information associated with the name and thus the name's referent. The scribe is not the referent, although he is (by hypothesis) the individual at the origin of a causal chain that began with the introduction of the name 'Melford Bibbins' into the language. It would be a mistake to suppose that reference can be explained here via associated descriptive content. Suppose that such content (*the proponent of such-and-such crazy philosophical views*) misconstrues the author's actual views but coincidentally describes precisely the views of some unknown hermit living somewhere in the Appalachian Mountains. In such a case, the author of the manuscript (the intuitive referent of the name) does not satisfy the associated content. Nevertheless, he is arguably the causal origin of that content insofar as he produced the work that generated it.

Devitt and Sterelny (1999) also advocated a hybrid theory. They claimed that reference fixing, to be successful, must involve a conceptual element. If you see a rock but think it is a tortoise and attempt to dub it 'Slow Poke,' reference fixing fails. They called this the 'qua-problem' and solved it by supposing that in order to fix the reference of a name, the speaker must have an accurate (if general) concept of the intended bearer.

Other Expressions

Philosophers often appeal to other expressions in developing their theories of proper names. They do so for (at least) two reasons: to clarify their theories and to provide methodological motivation for those theories. When a contrast is observed, clarification is generally the motivation. When a similarity is noted, promotion of the theory's generality is often the motivation.

Definite Descriptions

Definite descriptions are expressions of the form 'the F ,' such as 'the current vice president of the United

States.' How do these expressions compare with proper names? Millians generally argue they are importantly different: descriptions refer via satisfaction (or 'denotation'). Names refer directly, not in virtue of associated descriptive content. (For this reason, Millian theories are sometimes referred to as 'direct reference' theories.) Russell, in contrast, argued that proper names are 'disguised' definite descriptions and thus subject to the quantificational rendering provided by his theory of descriptions (1905). Frege (1892/1962) similarly treated definite descriptions as semantically of a piece with proper names, claiming that both sorts of expressions have their reference determined by their associated conceptual content (or 'sense').

Natural Kind Terms

Natural kind terms are terms that refer to kinds of things found in nature – 'water,' 'tiger,' and 'lemon' being stock examples. How do these expressions compare with proper names? Kripke (1980) suggested that these expressions are much like proper names: their reference is not determined by any associated conceptual content but by causal chains of the sort that link proper names to their bearers. These same ideas were developed by Putnam (1975) and later and in greater detail by Devitt and Sterelny (1999).

Bibliography

- Devitt M (1981). *Designation*. New York: Columbia University Press.
- Devitt M & Sterelny K (1999). *Language and reality*. Cambridge: MIT Press.
- Donnellan K (1970). 'Proper names and identifying descriptions.' *Synthese* 21, 335–358.
- Evans G (1973). 'The causal theory of names.' *Aristotelian Society Supplementary Volume* 47, 187–208.
- Evans G (1979). 'Reference and contingency.' *The Monist* 62, 161–189.
- Evans G (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Frege G (1962). 'On sense and reference.' In Black M & Geach P (eds.) *Philosophical writings*. Oxford: Blackwell. 56–78. (Original work published 1892.)
- Kripke S (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Marcus R B (1960). 'Extensionality.' *Mind* 69, 55–62.
- Marcus R B (1961). 'Modalities and intensional languages.' *Synthese* 13, 303–322.
- Mill J S (1843). *A system of logic*. London: Longman.
- Putnam H (1975). 'The meaning of "meaning."' In *Mind, language and reality*. Cambridge, UK: Cambridge University Press. 251–271.
- Recanati F (1993). *Direct reference: from language to thought*. Oxford: Blackwell.
- Russell B (1905). 'On denoting.' *Mind* 14, 479–493.

- Russell B (1917). 'Knowledge by acquaintance and knowledge by description.' In *Mysticism and logic*. London: Allen & Unwin. 152–167. (Original work published 1912.)
- Russell B (1919). *Introduction to mathematical philosophy*. London: Allen & Unwin.
- Russell B (1956). 'The philosophy of logical atomism.' In Marsh R (ed.) *Logic and knowledge*. London: Allen & Unwin. 177–281. (Original work published 1918.)

- Salmon N (1986). *Frege's puzzle*. Cambridge: MIT Press.
- Searle J (1958). 'Proper names.' *Mind* 67, 166–173.
- Searle J (1983). *Intentionality: an essay in the philosophy of mind*. New York: Cambridge University Press.
- Soames S (1987). 'Direct reference, propositional attitudes, and semantic content.' *Philosophical Topics* 15, 47–87.
- Wettstein H (2004). *The magic prism*. Oxford: Oxford University Press.

Propositional Attitude Ascription: Philosophical Aspects

L Clapp, Illinois Wesleyan University, Bloomington, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

Paradigmatic **propositional attitude ascriptions** (in English) are sentences of the form **A Vs that S**, where **A** is a singular definite noun phrase referring to a rational agent, **S** is a sentence, and **V** is a propositional attitude verb. Paradigmatic propositional attitude verbs include 'believe', 'doubt', 'realize', 'hope', 'regret', etc. Such verbs refer to **intentional mental states** with **representational content**, where this content can be evaluated for truth or falsity, and thus can be identified with a **proposition**. (What are sometimes called *verbs of saying*, e.g., 'say', 'deny', 'announce', are closely related to, but typically distinguished from, propositional attitude verbs.) Because propositional attitude verbs refer to intentional states with propositional content, it is natural to endorse a **relational analysis** of propositional attitude ascriptions: an assertion of 'John hopes that Mary won' says that the referent of 'John' bears the mental relation (or *attitude* in Russell's 1940 terminology) referred to by 'hopes' toward the proposition expressed by the complement clause 'that Mary won.' Similarly, an assertion of 'John doubts that Tom is happy' says that John bears a different relation, viz. doubting, toward a different proposition, viz. that Tom is happy.

Analytic philosophers have been, and continue to be, concerned to formulate an adequate semantic analysis of propositional attitude ascriptions for several reasons. One reason stems from philosophers' desire to understand the nature of mental states. What is, e.g., a belief? Is the existence of beliefs compatible with **physicalism**? If not, then should we be **eliminativists** about mental states, and deny that they really exist, or should we reject physicalism in favor of some sort of **dualism**? How can a mental state **represent** things as being a certain way? i.e., what is it for one entity to be **about** another? Does having a belief

require one to instantiate some sort of internal representational state (as **cognitivists** claim), or is it merely a matter of being disposed to behave in various ways (as **behaviorists** claim)? Whatever view one adopts toward these broader philosophical issues concerning the nature of mental states must be compatible with an adequate semantic analysis of how we talk about such states.

Another reason, perhaps the principle reason, philosophers are concerned with propositional attitude ascriptions stems from the close relationship between natural language and thought, and the apparent **compositionality** exhibited by both. What we can think is what is expressed by utterances of declarative sentences; if what we think are **thoughts**, and if what is expressed by utterances of declarative sentences are propositions, then what we think are propositions. Moreover, philosophers have argued that in order to explain the **productivity**, **interpretability**, and **systematicity** of language, and thus thought, some sort of **principle of semantic compositionality** must hold for language, and thus thought (see Frege (1914) and Davidson (1965) for appeals to productivity and interpretability. See Evans (1982) and Fodor (1998) for appeals to systematicity. Frege also appealed to compositionality to explain the **unity of the proposition**. See Frege (1891 and 1892b). Consequently, many analytic philosophers have endorsed some version of the following **principle of semantic compositionality**:

The proposition expressed by a sentence **S** in a context **c** is a function of (i) the **logical form** of **S**, and (ii) the **semantic values** invoked by **S** in **c**.

(The relativization to context is necessary to account for sentences containing **context sensitive** words.) The reason that philosophers interested in language and thought have been especially concerned with propositional attitude ascriptions is that these sentences provide apparent counterexamples against this otherwise well-motivated principle; this is known as the **opacity** problem for attitude ascriptions. The focus of my

remarks here will be on the opacity problem and influential proposals for its solution.

Consider the classic case of the hapless ancient astronomer: He used 'Hesperus' to refer to the first heavenly body to appear at dusk, and 'Phosphorus' to refer to the last heavenly body to disappear at dawn. But he did not realize that these names are coreferential; they both refer to Venus. Now consider the following attitude ascriptions:

- (1) The astronomer believed that Hesperus was visible at dusk.
- (2) The astronomer believed that Phosphorus was visible at dusk.

Given the astronomer's ignorance of the identity of Hesperus and Phosphorus, (1) is true and (2) is false. But that (1) and (2) differ in truth value is at least an apparent counterexample to the compositionality principle: If it is granted that (1) and (2) have the same logical form, and that the semantic value contributed by a word is its referent, then, given that 'Hesperus' and 'Phosphorus' are coreferential, semantic compositionality seems to require that (1) and (2) express the same proposition, and thus cannot differ in truth value. The opacity problem is sometimes called *substitution failure*: in (1) 'Phosphorus' cannot be substituted *salva veritate* for 'Hesperus.' Linguistic environments in which coreferential terms cannot be substituted *salva veritate* are known as **opaque**, as opposed to **transparent**, contexts.

Responses to the opacity problem can be divided into **conservative** proposals and **radical** proposals. Conservative proposals attempt to preserve the principle of semantic compositionality by somehow explaining away the apparent counterexamples. Conservative proposals can be subdivided into **semantic** conservative proposals, which appeal to the **meanings** of expressions to explain away the apparent counterexamples, and **pragmatic** conservative proposals, which appeal to how expressions are **used**. In contrast to conservative proposals, **radical proposals** accept the apparent counterexamples provided by attitude ascriptions, and thus reject the principle of semantic compositionality and offer alternative conceptions of semantics in its place.

To propose a **semantic conservative solution** to the opacity problem, then, is to provide an account of semantic values, and how they are invoked and combined, which preserves the above principle of compositionality yet explains why, e.g., (1) and (2) express different propositions. If the datum that (1) and (2) express different propositions is granted, and it is granted that (1) and (2) have the same logical form, then the general strategy one must adopt is clear: one must argue that (1) and (2) somehow invoke

additional, extrareferential, semantic values. Semantic conservative proposals all have this general strategy in common; where they differ is over **what** they take the requisite additional semantic values to be, and over **how** attitude ascriptions invoke different such additional semantic values.

The first, and probably most influential, semantic conservative proposal was presented by Frege (1892a). Indeed, Frege was the first to clearly formulate the opacity problem, and most subsequent work on the problem is a development of, or response to, Frege's seminal paper. A key feature of Frege's proposal is the thesis that words embedded in the complement clauses of attitude ascriptions **shift referents**; when words occur embedded in such **opaque** contexts, their semantic values are not their 'ordinary referents,' but are instead more finely individuated entities, which Frege called 'indirect referents.' Frege posited abstract objects he called **senses** to serve as indirect referents. A sense of a word is alleged to be a 'mode of presentation' of the referent of the word (if the word has a referent). To return to the hapless astronomer, Frege would say that the referent, Venus, is presented to the astronomer in two different ways; one of these ways is the sense of 'Hesperus,' and the other is the sense of 'Phosphorus.' The astronomer fails to recognize that Hesperus is Phosphorus because he fails to realize that these different modes of presentation are in fact modes of presentation of, or ways of thinking about, the same referent. If it is granted that 'Hesperus' and 'Phosphorus,' though coreferential, nonetheless express different senses, then, because of referent shifting, (1) and (2) invoke different semantic values. And consequently, that these sentences express different propositions, and have different truth values, is no threat to semantic compositionality.

Many theorists endorse Frege's thesis that words embedded in the complement clauses of attitude ascriptions shift referents, yet, influenced by broader philosophical commitments, they reject senses, the abstract objects Frege posited to serve as indirect referents. Other sorts of entities posited to serve as indirect referents include **intensions** (Montague, 1974), sentences (Carnap, 1946), and **mental representations** (Fodor, 1978). All theories which allow for referent shifting, however, face a common difficulty: As Quine (1956) noted, attitude ascriptions are ambiguous (or at least context sensitive). For some utterances of attitude ascriptions, what is relevant for determining truth conditions is simply the referent of an embedded word, rather than the word, or some other sort of indirect referent. And for these so-called *de re* utterances of attitude ascriptions, Frege's referent shifting thesis is inappropriate. But, as Frege

noted, for other, so-called *de dicto*, utterances of attitude ascriptions, it seems that something other than the referents of embedded words is relevant for determining truth conditions. It is now widely accepted that in different contexts the same attitude ascription (sentence type) can express different truth conditions; in particular, some utterances of, e.g., the type of (1) are *de re* (or **transparent**), while other utterances like (1) are *de dicto* (or **opaque**).

The recognition of the context sensitivity of attitude ascriptions has led philosophers to add complexities to the semantic conservative strategy. Quine (1956) proposes that, if propositional attitudes are to be countenanced at all, propositional attitude verbs ought to be treated as ambiguous between a **relational** (*de re*) sense, and a **notional** (*de dicto*) sense. More recent proposals build on Carnap's (1946) theory, and identify indirect referents not with words, but with abstract structures composed of both words and their ordinary referents (relative to a context) combined. These Carnap-inspired theories attempt to account for the context sensitivity of attitude ascriptions by allowing what is necessary for an agent to hold the relevant mental relation to the combined entity to vary across contexts (see Richard, 1990 and Higginbotham, 1991).

Thus far, I have reviewed semantic conservative proposals that follow Frege in maintaining that words in complement clauses of attitude ascriptions shift their referents; where the proposals discussed thus far have differed from each other concerns **what** the additional semantic values are. Other semantic conservative proposals, however, disagree with Frege not only concerning **what** the requisite additional semantic values are, but they also reject Frege's referent shifting thesis. An early alternative to Frege's indirect reference strategy was proposed by Russell (1919). Russell's proposal preserves the simple idea that semantic values are **always only** ordinary referents, but he achieves this result at the cost of allowing the **logical form** of a sentence to differ significantly from its **surface form**. According to Russell, what appear to be referring terms, e.g., 'Hesperus' and 'Phosphorus,' are merely abbreviations for "denoting phrases" composed of quantifiers and unpronounced genuine referring terms; Russell called these posited unpronounced genuine referring terms "logically proper names." According to Russell then, despite appearances, Venus is not the referent of any genuine referring term in the logical form of either (1) or (2). Rather, the apparent names 'Hesperus' and 'Phosphorus' abbreviate denoting phrases comprised of different, noncoreferential, logically proper names. Thus, that (1) and (2) express different propositions is not a counterexample to semantic compositionality,

because, according to Russell, the sentences invoke different semantic values. (Not surprisingly, Russell never presents a full reduction of an apparent name into a denoting phrase comprised of logically proper names.)

A more recent proposal similar in spirit to Russell's is the "hidden indexical" analysis (see Schiffer, 1977 and Crimmins, 1992). This proposal maintains that the additional semantic values required for distinguishing the propositions expressed by utterances of (1) and (2) are the referents of 'hidden' – phonologically unrealized – indexical items that are somehow present in the logical form, but not in the surface form, of the sentence. Crimmins takes the posited hidden indexicals to refer (in contexts) to "mental particulars," token mental representations that, similar to Frege's senses, are ways of thinking of ordinary referents. The hidden indexical analysis thus does not deny that, e.g., Venus is a semantic value invoked by both utterances (1) and (2), because Venus is the referent of both 'Hesperus' and 'Phosphorus.' But that such utterances express different propositions is nonetheless not a counterexample to semantic compositionality because the presence of hidden indexicals allows the utterances to invoke different mental particulars as 'unarticulated' semantic values.

Another influential semantic conservative proposal that rejects referent shifting is Davidson's (1968) "paratactic" analysis of indirect speech. Davidson proposes that an indirect speech report with the surface form 'A said that S' has the logical form of **two** sentences: first, 'A said **that**,' and second, 'S.' The word 'that' in the first sentence is interpreted as a **demonstrative**, and in an utterance with the surface form 'A said that S,' the demonstrative refers to the subsequent **utterance** of 'S.' Thus, in Davidson's proposal, the additional semantic value invoked by an utterance of an indirect speech report is the very act of uttering the sentence embedded in the complement clause. According to Davidson's proposal, an utterance with the surface form 'A said that S' is true just in the case the agent referred to by 'A' performed an utterance that is relevantly similar to the referent of 'that,' where this referent is the **utterance** of the embedded 'S.' Davidson applied his analysis only to indirect speech reports, but others, such as Lepore and Loewer (1989), have extended Davidson's analysis so that it applies to attitude ascriptions.

Whereas semantic conservative proposals accept that, e.g., (1) and (2) express different propositions and attempt to reconcile this datum with the principle of semantic compositionality, **pragmatic conservative proposals** reject the datum and thereby undermine the apparent counterexample to semantic compositionality. According to pragmatic conservative

proposals, (1) and (2) do not present a counterexample to compositionality because, despite our pre-theoretic judgments to the contrary, they do not express different propositions, and they do not differ in truth value. Support for this seemingly implausible strategy comes from two principle sources: First, there are influential arguments due to Kripke (1972) and Kaplan (1977) in support of the **thesis of direct reference**, according to which the only semantically relevant feature ever associated with some terms – names, demonstratives, and indexicals being the paradigmatic cases – is the referent of the term. The thesis of direct reference thus entails that, e.g., utterances (1) and (2) express the same proposition. Second, there is Grice's (1975) distinction between **what is said** by an utterance and **what is merely implied** by an utterance. According to Grice, **what is said** is the domain of **semantics** and thus must conform to the compositionality principle, while **what is implied** is the domain of **pragmatics**, and thus need not conform to the principle. According to the pragmatic conservative strategy then, utterances (1) and (2) semantically say the same thing, but they pragmatically imply different things. And our pre-theoretic judgment that utterances (1) and (2) say different things is a result of our naively conflating **semantically expressed** information with **pragmatically implied** information. Pragmatic conservative proposals are developed by Salmon (1986) and Soames (2002).

In contrast to **conservative proposals** which attempt to preserve semantic compositionality by somehow explaining away the at least apparent counterexamples posed by attitude ascriptions, **radical proposals** accept the counterexamples and reject semantic compositionality. Given the fundamental role that the principle of compositionality has played in semantics and analytic philosophy generally, a theorist who endorses a radical proposal must provide an alternative conception of what it is to provide a semantic analysis. Moreover, this alternative conception must be at least compatible with the phenomena of **productivity**, **interpretability**, and **systematicity**. One sort of radical proposal falls within the broader movement known as **Radical Pragmatics**. Radical Pragmatics is characterized by the thesis that the linguistic meaning associated with an utterance underdetermines what is said by the utterance; i.e., knowing the logical form of an utterance and knowing the referents of all the words (or other relevant features) does not suffice for determining what is said by the utterance; in order to arrive at something truth-evaluable, additional processing is required. **Relevance Theory**, as proposed by Sperber and Wilson (1986), is a paradigmatic example of Radical Pragmatics. Thus, from the perspective of Radical Pragmatics, it is not

particularly problematic that utterances (1) and (2) express distinct truth conditions. A version of Radical Pragmatics is applied to attitude ascriptions in Bach (1997).

Another sort of radical proposal falls within the broader movement known as **Dynamic Semantics**. According to Dynamic Semantics, the semantic content of a declarative utterance is not a truth-evaluable proposition, but rather linguistically encoded general instructions (i.e., a 'context change potential') for updating the set of beliefs shared between the speaker and his audience (i.e., the 'common ground'). To determine **what is said** by an utterance, one must apply the linguistically encoded general instructions to the set of shared beliefs. Thus, **what is said** depends not only upon the logical form of the uttered sentence and the semantic values of the words in it, but also upon what the shared beliefs of the speaker and his audience happen to be. A paradigmatic example of Dynamic Semantics is **Discourse Representation Theory**, as proposed by Kamp and Ryle (1993). Because utterances (1) and (2) encode different instructions for updating sets of shared beliefs, **what is said** by such utterances (relative to the same set of shared beliefs) can differ. Thus, from the perspective of Dynamic Semantics also, it is not particularly problematic that such utterances are interpreted as expressing distinct truth conditions. This sort of dynamic approach is applied to the phenomenon of opacity in Asher (1993).

See also: Indexicality: Philosophical Aspects; Metaphysics, Substitution *Salva Veritate* and the Slingshot Argument; Semantic Value.

Bibliography

- Asher N (1993). *Reference to abstract objects in discourse*. Dordrecht: Kluwer.
- Bach K (1997). 'Do belief reports report beliefs?' *Pacific Philosophical Quarterly* 78, 215–241.
- Beaney M (ed.) (1997). *The Frege reader*. Oxford: Blackwell.
- Carnap R (1946). *Meaning and necessity*. Chicago: University of Chicago Press.
- Crimmins M (1992). *Talk about beliefs*. Cambridge: MIT Press.
- Davidson D (1965). 'Theories of meaning and learnable languages.' In Davidson (ed.) (1984), 3–15.
- Davidson D (1968). 'On saying that.' *Synthese* 19, 158–174. Repr. In Davidson (ed.) (1984), 93–108.
- Davidson D (1984). *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
- Evans G (1982). McDowell J (ed.) *The varieties of reference*. Oxford: Oxford University Press.

- Fodor J (1978). 'Propositional attitudes.' *The Monist* 64. Repr. In Fodor J (ed.) (1981). *Representations*. Cambridge: MIT Press. 177–203.
- Fodor J (1998). *Concepts: where cognitive science went wrong*. Oxford: Oxford University Press.
- Frege G (1891). 'Function and concept.' In Beaney M (ed.). 130–148.
- Frege G (1892a). 'On *Sinn* and *Bedeutung*.' In Beaney M (ed.). 172–180.
- Frege G (1892b). 'On concept and object.' In Beaney M (ed.). 181–193.
- Frege G (1979). 'Logic in mathematics.' In Hermes H, Kambartel F & Kaulbach F (eds.) *Posthumous writings*. Chicago: University of Chicago Press. 203–250.
- Grice H P (1975). 'Logic and conversation.' In Cole P & Morgan J (eds.) *Syntax and Semantics 3: Speech Acts*. New York: Academic Press. 41–58. Repr. In Grice H P (ed.) *Studies in the way of words*. Cambridge: Harvard University Press. 41–58.
- Higginbotham J (1991). 'Belief and logical form.' *Mind and Language* 6, 344–369.
- Kamp H & Ryle U (1993). *From discourse to logic*. Dordrecht: Kluwer.
- Kaplan D (1977). 'Demonstratives draft #2.' In Almog J, Perry J & Wettstein H (eds.) (1989) *Themes from Kaplan*. Oxford: Oxford University Press. 481–563.
- Kripke S (1972). 'Naming and necessity.' In Davidson D & Harman G (eds.) *Semantics of natural language*. Dordrecht: Reidel Publishing. 253–255.
- Lepore E & Loewer B (1989). 'You can say *that* again.' In French P, Uehling T & Wettstein H (eds.) *Midwest studies in philosophy* 14, 338–356.
- Montague R (1974). *Formal philosophy*. New Haven: Yale University Press.
- Quine W V O (1956). 'Quantifiers and propositional attitudes.' *The Journal of Philosophy* 53, 177–187.
- Richard M (1990). *Propositional attitudes*. Cambridge: Cambridge University Press.
- Russell B (1919). *Introduction to mathematical philosophy*. London: George Allen and Unwin.
- Russell B (1940). *An inquiry into meaning and truth*. New York: Norton.
- Salmon N (1986). *Frege's puzzle*. Cambridge: MIT Press.
- Schiffer S (1977). 'Naming and knowing.' In French P, Uehling T & Wettstein H (eds.) *Midwest studies in philosophy* 2, 28–41.
- Soames S (2002). *Beyond rigidity: the unfinished semantic agenda of naming and necessity*. Oxford: Oxford University Press.
- Sperber D & Wilson D (1986). *Relevance: communication and cognition*. Oxford: Blackwell.

Propositions

S Crawford, Lancaster University, Lancaster, UK

© 2006 Elsevier Ltd. All rights reserved.

When an English speaker utters the sentence 'Snow is white' and a French speaker utters the sentence 'La neige est blanche,' there is a clear sense in which they have both said the same thing. Moreover, given that they both intend sincerely to express their beliefs by uttering these sentences, there seems a clear sense in which they both believe the same thing. Philosophers call this thing that both have said and that both believe a *proposition*. The declarative sentences we utter and write down are said to have propositions as their linguistic meanings and to express those propositions. Two or more sentences (in the same language or in different ones) can have the same proposition as their meaning – that is, they can express the same proposition – just as different numerals (e.g., the Arabic numeral '4' and the Roman numeral 'IV') can designate the same number, namely four. Moreover, propositions are supposed to be the contents of many of our mental states, such as belief, knowledge, doubt, supposition, memory, desire, intention, and so on. Russell (1903) called these mental states "propositional attitudes."

Roles Played by Propositions

The notion of a proposition has played, and continues to play, important roles in the philosophy of logic, philosophy of language, and philosophy of mind. Three of those roles have been mentioned already: Propositions are the linguistic meanings of the sentences we utter or write down, they are the contents of our sayings when we utter or write down those sentences, and they are the contents of our thoughts. Propositions are commonly thought to have other important functions as well. They are often taken to be the "primary bearers" of truth and falsity, in the sense that the sentences we utter and the beliefs we have are true or false only derivatively in virtue of the propositions associated with them being true or false. Propositions are also said to be the primary bearers of modal properties, such as necessity, contingency, possibility, impossibility, and so on. And logical relations, such as consistency, inconsistency, and entailment, are said to hold between or among propositions.

It is a good question whether any single notion of a proposition can play all these roles at once. Sentences containing indexicals, for example, cause difficulties in this regard, for they seem to suggest that the linguistic meaning of a sentence may differ from the

proposition it expresses. When I say ‘I’m hungry,’ and when you say ‘I’m hungry,’ there is a sense in which our respective sentences have the same meaning; but it also seems clear that they express different propositions, one about me and the other about you. Conversely, when I say ‘I’m hungry’ and you say, addressing me, ‘You’re hungry,’ we seem to have said the same thing – that is, expressed the same proposition, a proposition about me – but uttered sentences with different meanings. In light of this, Kaplan (1989) has distinguished between the *character* (roughly, linguistic meaning) and the *content* (proposition expressed) of a sentence. Whether any single notion of a proposition can play all these different roles depends very much on what propositions are taken to be. So what then are propositions exactly? What are these entities, propositions, that we assert and believe and that are true or false and necessary or contingent?

Propositions as Abstract Entities

The history and development of the concept of the proposition are a long and complex story (Gale, 1967); but contemporary accounts derive most immediately from Frege’s attack on psychologism and the *fin de siècle* revolt against idealism inaugurated by Russell and Moore (Frege, 1892a, 1918; Moore, 1899; Russell, 1903; Hylton, 1984). The core of this shared account is the idea that propositions are mind-independent, extralinguistic abstract entities akin to numbers, mathematical functions, and sets.

This core idea subsequently came under persistent attack by Quine (1960, 1986) who viewed propositions (and other ‘intensional’ entities such as properties and relations) as ‘creatures of darkness’ owing to the alleged lack of any criteria for individuating them, as well as their essential involvement with what Quine (1951) deemed the obscure and suspect notions of meaning and synonymy. In the service of his sparse desert ontology, Quine proposed to replace abstract propositions with concrete sentences (propositions are really just unnecessary ‘shadows’ of sentences anyway, to invoke Wittgenstein’s metaphor) or, more accurately, with mathematical sequences of word-tokens that have been uttered or inscribed at some time (Quine, 1960). However, neither Quine’s relentless critique of propositions as abstract entities nor his replacement of them with sentences has been widely accepted. Moreover, despite Quine’s closely related attack on modal notions, such as necessity and possibility, the rise of modal logic and modal metaphysics continued unabated and accorded a

central place to propositions construed as abstract entities. The seminal work of Kripke (1972) was instrumental in turning the tide against Quine.

Let us then turn to the most popular contemporary approaches to the nature of propositions as mind-independent, extralinguistic abstract entities. As we shall see, there has been some lack of clarity about whether theories of propositions are theories about what propositions *are*, ontologically speaking, or whether theories of propositions are theories about how they should best be represented or modeled in a formal semantic theory.

Two Approaches: Structured and Structureless Entities

Generally speaking, theories of propositions can be divided into two sorts. The first takes propositions to be certain kinds of structured entities with components standing in various relations to each other, with the structure of the entities roughly mirroring the structure of the sentences that express them. The second takes propositions to be structureless entities. The structureless approach invokes the notion of a “possible world,” borrowed from the new model theory for modal logic developed in the 1950s and 1960s, known as “possible worlds semantics.” A possible world is a “way things could have been.” The idea is that a proposition is a set of possible worlds (the set of worlds at which the proposition is true), or equivalently, a proposition is a function (in the mathematical sense) from the set of possible worlds to the set of true values.

That the possible-worlds approach to propositions treats propositions as structureless is easily seen. Consider the sentences ‘Russell is not alive and Frege is not alive’ and ‘It is not true that either Russell is alive or Frege is alive.’ These sentences have very different structures: the first is a conjunction of two negations, and the second is a negation of a disjunction. Yet, they are true in the same possible worlds because they are logically equivalent. According to the possible-worlds approach, both sentences therefore express the same proposition.

According to the structured approach, however, these two sentences express different propositions because in the structured approach a proposition is not individuated in terms of the possible worlds in which it is true, but rather in terms of the constituents that make up the proposition. On this account, a proposition is a structured entity composed of parts, where the parts and the order of composition more or less mirror the parts and order of composition of the sentences that express the proposition.

Two Structured Approaches: Russellian and Fregean

There are, broadly speaking, two types of structured theory, the first deriving from Russell (1903) and the second from Frege (1892a), which are distinguished from each other by what they take the constituents of propositions to be. According to the Russellian (or “neo-Russellian”) approach, the constituents of the propositions we assert and believe, and to which truth and falsity and modal properties belong, are the objects, properties, and relations that our assertions and beliefs are about: desks, trees, other people, and other everyday objects. According to the Fregean approach, propositional constituents are rather ‘senses’ or ‘modes of presentation’ (or concepts or ‘ways of thinking’) of the objects, properties, and relations that our saying and thoughts are about: modes of presentation of desks, trees, and other people. There are also combination views, according to which propositional constituents are both the things our sayings and thoughts are about and modes of presentation of those things.

On the Russellian view, propositions are identified with ordered pairs of n objects and an n -place relation: $\langle \langle x_1, \dots, x_n \rangle, X^n \rangle$. The proposition that Maggie is cooking is thus identical with the ordered pair: $\langle \text{Maggie}, \text{cooking} \rangle$. The Fregean view identifies propositions with ordered pairs of n modes of presentation (of objects) and an n -place mode of presentation (of a relation): $\langle \langle m_1, \dots, m_n \rangle, M^n \rangle$. Here, m_1, \dots, m_n are modes of presentation of x_1, \dots, x_n , and M^n is a mode of presentation of X^n . The proposition that Maggie is cooking, for the Fregean, is thus the ordered pair: $\langle \text{mode of presentation of Maggie}, \text{mode of presentation of cooking} \rangle$.

A third view, which combines Russellian and Fregean elements, has it that the propositions we believe contain both the items our beliefs are about and modes of presentation. For example, we might believe or assert ‘quasisingular propositions’ (Schiffer, 1978; Recanati, 1993) that contain objects, properties, and relations, as well as modes of presentation of all these things: $\langle \langle \langle x_1, m_1 \rangle \dots \langle x_n, m_n \rangle \rangle, \langle X^n, M^n \rangle \rangle$. According to this view, when Tom believes Maggie is cooking, the proposition he believes is identical with the ordered pair: $\langle \langle \text{Maggie}, \text{mode of presentation of Maggie} \rangle, \langle \text{cooking}, \text{mode of presentation of cooking} \rangle \rangle$. It is possible, of course, to hold that sometimes we believe one type of proposition (e.g., a Russellian one) and sometimes we believe another type (e.g., a quasisingular one), depending on the situation and context in question. For present purposes, however, the important point is that propositional constituents are either objects, properties, and

relations or modes of presentation of these things (often just called *concepts* of them – though this contemporary use of the term ‘concept’ should be distinguished from Frege’s very different use of that term). And the propositions we believe – if there are such things – are composed out of such items in some way or another.

Ontology or Semantics?

So far, we have been speaking as if both the structureless and structured accounts of propositions were accounts of what propositions are. If this is correct, then, on the structureless account, what we assert when we assert a proposition and what we believe when we believe a proposition is a set or a function. Similarly, according to the structured accounts, to assert or to believe that Maggie is cooking is to stand in a relation to an ordered set. On the face of it, this does indeed seem very implausible (Bealer, 1998). Moreover, it is hard to understand how ordered sets could be true or false or necessary or contingent. These set-theoretic constructions just do not seem to be the kinds of things that can have the properties that propositions are supposed to have.

The structured position suffers from another problem too: *which* ordered set is the proposition that Maggie is cooking? Taking the Russellian approach as an illustration, is it $\langle \text{Maggie}, \text{cooking} \rangle$ or is it $\langle \text{cooking}, \text{Maggie} \rangle$? There seems no way to determine (nonarbitrarily) which ordered set a certain proposition is (Bealer, 1998; Jubien, 2001). This general difficulty of reducing abstract objects, such as numbers, to sets is due to Benacerraff (1965).

In light of this, it seems best to take the structureless and structured approaches as rival ways of representing propositions in a formal semantic theory and not as accounts of what propositions are. Whether or not this is an adequate response to the foregoing objections, it is evident that, even interpreted as proposals about how to represent propositions, each approach is still not without its problems.

A Problem for the Structureless Approach

The structureless account appears to face what seems to be a devastating objection, stemming from the fact that it implies that all necessarily equivalent propositions are identical. Since the sentences ‘A sister is a female sibling’ and ‘A brother is a male sibling’ are true in all possible worlds, they express the same proposition. But, these sentences obviously have different meanings and thus seem to express different propositions. Moreover, if belief is a binary relation between a subject and a proposition, and a

proposition is a set of possible worlds, then, for any necessarily equivalent propositions *P* and *Q*, if *S* believes *P*, then it follows that *S* believes *Q*. So, for example, if Maggie believes that $5 + 5 = 10$, she must then believe that arithmetic is incomplete, since both these things are true in every possible world. Valiant efforts have been made to mitigate these counter-intuitive consequences (Stalnaker, 1984), but none has achieved widespread acceptance.

The very structurelessness of propositions on the possible worlds account means that propositions are not “fine grained” enough to serve as the objects of attitudes and the meanings of sentences. There are just not enough to go around on the possible worlds account. (Another approach to this problem, within the possible worlds framework, found in the work of Lewis (1972) and Montague (1974), invokes “structured intensions” and derived from Carnap’s (1956) notion of “intensional isomorphism types.”)

Problems for the Structured Approaches

The majority of philosophers, then, adopt a structured approach, either a Russellian one or a Fregean one. Turning first to the Russellian representation of propositions, it is evident that it is able to distinguish among necessarily equivalent propositions. The propositions expressed by the sentences ‘Sisters are female siblings’ and ‘Brothers are male siblings’ have different constituents and are therefore distinct. So the Russellian manages to achieve a certain fineness of grain in its representation of propositions. It is a matter of controversy, however, whether his representation of propositions is fine enough. For the Russellian, the sentences ‘George Eliot is a novelist’ and ‘Mary Anne Evans is a novelist’ express the same proposition, because the parts of each sentence (e.g., the proper names and predicates) refer to (and are true of) the same things, and it is these same things that form the constituents of the propositions that each expresses. Now, whether or not we want to say that these sentences have different meanings, it has seemed to some philosophers that we do want to say that someone could believe George Eliot is a novelist without also believing that Mary Ann Evans is a novelist (even though George Eliot is Mary Ann Evans). In other words, belief content – and cognitive content more generally – seems to be extremely fine-grained. Importantly, cognitive content seems to be more fine-grained than a representation – using only the worldly objects, properties, and relations that our thoughts are about, can cope with. Since, for the Russellian, the two aforementioned sentences express the same proposition, and belief for the Russellian is a relation between a subject and a proposition,

it follows that it is impossible for the two sentences ‘Ralph believes that George Eliot is a novelist’ and ‘Ralph believes that Mary Ann Evans is a novelist’ to differ in truth-value. Philosophers of a Fregean bent, however, argue that they can differ in truth-value – that such a difference in truth-value is required, for example, to explain the linguistic and nonlinguistic behaviour of agents – and that the Russellian theory must therefore be wrong. Russellians have offered systematic replies to these arguments (Salmon, 1986; Soames, 1987; Braun, 1998) but the debate remains open.

The Fregean conception of propositions exhibits perhaps the finest grain and is thus able both to distinguish among necessarily equivalent propositions (like the Russellian account but unlike the unstructured possible world account) and to allow for a difference in truth-value between the two foregoing belief sentences. Because the constituents of propositions are modes of presentation (or concepts or “senses”) of the things that our beliefs are about, and modes of presentation are many-one related to these things, there will be many more modes of presentation than the things they are modes of presentation of; and so the requisite fineness of grain necessary for distinguishing between the propositions expressed by the sentences ‘George Eliot is a novelist’ and ‘Mary Anne Evans is a novelist’ will be achieved. Since these two sentences express different propositions due to their different conceptual constituents, it is possible for a person to believe the proposition expressed by one but not to believe the proposition expressed by the other.

This extra fineness of grain is bought at a certain cost, however, for it seems that the propositional grain is perhaps now too fine for propositions to play the role of the linguistic meanings of sentences. Arguably, since, for example, ‘attorney’ and ‘lawyer’ are synonyms, the sentences ‘Lawyers are wealthy’ and ‘Attorneys are wealthy’ are synonymous, that is, have the same meaning. They should therefore express the same propositions (if propositions are supposed to be the meanings of sentences). Yet, it seems possible for a person to believe that lawyers are wealthy, but to doubt whether attorneys are, indicating for the Fregean that the two sentences express different propositions. If the linguistic meanings of sentences are more coarsely grained than the cognitive contents of thoughts according to the Fregean, it is not clear whether a single thing – a proposition – can play both roles.

The Fregean theory is also incomplete in a way that the Russellian theory is not, and this takes us back to the ontological question of what propositions *are*, as opposed to the question of how best to represent

them. The Russellian has a clear account of what the constituents of propositions are: the ordinary objects, properties, and relations that our thoughts and sayings are about. For the Fregean, however, the constituents of propositions are sense or modes of presentation or concepts of the things that our thoughts and sayings are about. Yet, what are these modes of presentation, these senses? All attempts to say what they are have been subjected to powerful and sustained criticism, and for this reason it is not clear whether the Fregean theory can ever be completed (Schiffer, 2003).

However, just because the Russellian has an account of what propositional constituents are does not mean that the Russellian theory is complete. For the Russellian theory suffers from the lack of any account of what binds propositional constituents together to form a unity, rather than a loose collection of unrelated parts. How is a proposition distinguished from a mere *list* of items? Frege (1892b) himself held that the key to this question was to be found in the nature of the propositional constituents themselves: At least one of them is always “incomplete” or “unsaturated” and is “completed” or “saturated” by the other constituents, which are themselves already complete or saturated. Russell (1913) struggled heroically for many years, indeed decades, to solve this “binding” problem or “the problem of the unity of the proposition” (a problem first stated by Plato in *The sophist*) without ultimate success. Contemporary Russellians, however, have been more concerned with the construction of formal semantic theories that invoke model-theoretic representations of propositions than with the metaphysical question of what binds or glues propositional constituents together into a propositional unity. This difficult and ancient question is once again beginning to receive the attention it deserves (Gaskin, 1995; King, 1995; Jubien, 2001; Gibson, 2004).

See also: Character versus Content; *De Dicto* versus *De Re*; Intention and Semantics; Modal Logic; Objects, Properties, and Functions; Propositional Attitude Ascription; Philosophical Aspects; Sense and Reference; Philosophical Aspects; Truth: Primary Bearers.

Bibliography

- Bealer G (1998). ‘Propositions.’ *Mind* 107(425), 1–32.
 Beaney M (ed.) (1997). *The Frege reader*. Oxford: Basil Blackwell.
 Benacerraf P (1965). ‘What numbers could not be.’ *Philosophical Review* 74, 47–73.
 Braun D (1998). ‘Understanding belief reports.’ *Philosophical Review* 107, 555–595.

- Carnap R (1956). *Meaning and necessity* (2nd edn.). Chicago: University of Chicago Press.
 Cartwright R (1962). ‘Propositions.’ In Butler R J (ed.) *Analytical philosophy*, vol. 1. Oxford: Basil Blackwell.
 Davidson D & Harman G (eds.) (1972). *Semantics of natural language*. Dordrecht: D. Reidel.
 Evans G (1982). *The varieties of reference*. McDowell J (ed.). Oxford: Clarendon Press.
 Frege G (1892a). ‘On Sinn and Bedeutung,’ Black M (trans.). In Beaney (ed.) (1997).
 Frege G (1892b). ‘On Concept and object,’ Geach P (trans.). In Beaney (ed.) (1997).
 Frege G (1914). ‘Letter to Jourdain,’ Kaal H (trans.). In Beaney (ed.) (1997).
 Frege G (1918). ‘Thought,’ Geach P & Stoothoff R H (trans.). In Beaney (ed.) (1997).
 Gale R (1967). ‘Propositions, judgments, sentences and statements.’ In Edwards P (ed.) *The encyclopedia of philosophy*. New York: Macmillan.
 Gaskin R (1995). ‘Bradley’s regress, the copula and the unity of the proposition.’ *Philosophical Quarterly* 45(179), 161–180.
 Gibson MI (2004). *From naming to saying. The unity of the proposition*. Oxford: Basil Blackwell.
 Hylton P (1984). ‘The nature of the proposition and the revolt against idealism.’ In Rorty R, Schneewind J B & Skinner Q (eds.) *Philosophy in history*. Cambridge: Cambridge University Press.
 Jubien M (2001). ‘Propositions and the objects of thought.’ *Philosophical Studies* 104, 47–62.
 Kaplan D (1989). ‘Demonstratives.’ In Almog J, Perry J & Wettstein H (eds.) *Themes from Kaplan*. Oxford: Oxford University Press.
 King J C (1995). ‘Structured propositions and complex predicates.’ *Nous* 29(4), 516–535.
 Kripke S (1972). ‘Naming and necessity.’ In Davidson D & Harman G (eds.) Reprinted as *Naming and necessity* (1980). Cambridge, MA: Harvard University Press.
 Lewis D (1972). ‘General semantics.’ In Davidson & Harman (eds.).
 Montague R (1974). In Thomason R (ed.) *Formal Philosophy*. New Haven: Yale University Press.
 Moore G E (1899). ‘The nature of judgement.’ *Mind* 8, 176–193.
 Quine W V O (1951). ‘Two dogmas of empiricism.’ In Quine W V O (1953) *From a logical point of view*. Cambridge, MA: Harvard University Press.
 Quine W V O (1960). *Word and object*. Cambridge, MA: MIT Press.
 Quine W V O (1986b). *Philosophy of logic* (2nd edn.). Cambridge, MA: Harvard University Press.
 Recanati F (1993). *Direct reference*. Oxford: Basil Blackwell.
 Richard M (1990). *Propositional attitudes*. Cambridge: Cambridge University Press.
 Russell B (1903). *The principles of mathematics*. London: Routledge.
 Russell B (1913). *Theory of knowledge*. Reprinted London: Routledge (1992).
 Salmon N (1986). *Frege’s puzzle*. Cambridge, MA: MIT Press.

Salmon N & Soames S (eds.) (1988). *Propositions and attitudes*. Oxford: Oxford University Press.

Schiffer S (1978). 'The basis of reference.' *Erkenntnis* 13, 171–206.

Schiffer S (2003). *The things we mean*. Oxford: Clarendon Press.

Soames S (1987). 'Direct reference, propositional attitudes, and semantic content.' *Philosophical Topics* 15.

Stalnaker R (1976). 'Propositions.' In McKay & Lerrill (eds.) *Issues in the philosophy of language*. New Haven: Yale University Press.

Stalnaker R (1984). *Inquiry*. Cambridge, MA: MIT Press.



Quantifiers: Semantics

E L Keenan, University of California, Los Angeles, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

During the past 25 years, our empirical and mathematical knowledge of quantification in natural language has exploded. We now have mathematically precise (if sometimes contentious) answers to questions raised independently within generative grammar, and we are able to offer many new generalizations. We review these results here. For extensive overviews, see Westerståhl (1989), Keenan (1996), and Keenan and Westerståhl (1997). Some important collections of articles are van Benthem and ter Meulen (1985), Reuland and ter Meulen (1987), Gärdenfors (1987), Lappin (1988), van der Does and van Eijck (1996), Szabolcsi (1997), and from a mathematical perspective, Krynicki *et al.* (1995).

The best understood type of quantification in natural language is that exemplified by *all poets* in *All poets daydream*. We treat *daydream* as denoting a property of individuals, represented as a subset of the domain E of objects under discussion. Quantified Noun Phrases (NPs) such as *all poets* will be treated as denoting functions, called ‘generalized quantifiers,’ which map properties to truth values, True (T) or False (F). For example, writing denotations in bold-face, **all poets** maps **daydream** to True (over a domain E) if and only if the set of poets is a subset of the set of objects that daydream. Interpreting Dets (Determiners) as functions from properties to generalized quantifiers, we give denotations for many quantifiers in simple set theoretical terms. We write $X \subseteq Y$ to say “X is a subset of Y,” $X \cap Y$ for “X intersect Y,” the set of objects that lie in both X and Y; $X - Y$ for the set of objects in X that are not in Y, and $|X|$ for the number of elements in X.

- (1a) **ALL**(A) (B) = T iff $A \subseteq B$
- (1b) **(THE TEN)** (A) (B) = T iff $|A| = 10$ and $A \subseteq B$
- (1c) **NO**(A) (B) = T iff $|A \cap B| = 0$
- (1d) **SOME**(A) (B) = T iff $|A \cap B| > 0$
- (1e) **NEITHER**(A) (B) = T iff $|A| = 2$ & $|A \cap B| = 0$

$$(1f) \text{ (FEWER THAN TEN) (A) (B) = T iff } |A \cap B| < 10$$

$$(1g) \text{ (ALL BUT ONE) (A) (B) = T iff } |A - B| = 1$$

$$(1h) \text{ MOST(A) (B) = T iff } |A \cap B| > |A - B|$$

To test that the definitions above have been properly understood, the reader should try to fill in appropriately the blanks in (2).

$$(2a) \text{ (AT LEAST TWO) (A) (B) = T iff } \underline{\hspace{2cm}}$$

$$(2b) \text{ BOTH(A) (B) = T iff } \underline{\hspace{2cm}}$$

$$(2c) \text{ (AT MOST FIVE OF THE TEN) (A) (B) = T iff } \underline{\hspace{2cm}}$$

$$(2d) \text{ (LESS THAN HALF THE) (A) (B) = T iff } \underline{\hspace{2cm}}$$

We concentrate on NPs of the form [Det + N], but we should point out three further classes of generalized quantifiers denoting NPs: first, lexical NPs, notably proper nouns such as *John* and *Mary*; second, boolean compounds in *and*, *or*, *neither ... nor ...* and *not*, as in *Neither John nor Mary* (came to the party), *Sue and some student* set up the chairs, *not more than two students* attended the lecture. And third, NPs built from Dets that combine with two Nouns to form an NP (Keenan and Moss, 1985; Beghelli, 1993): *more students than teachers* (signed the petition), *exactly as many students as teachers* signed, etc. The interpretation of *more ... than ...* is given by:

$$(3) \text{ For all properties A,B,C (MORE A THAN B) (C) = T iff } |A \cap C| > |B \cap C|$$

The reader may want to write out the definitions of two place Dets such as **FEWER ... THAN ...**, **TWO MORE ... THAN ...**, **TWICE AS MANY ... AS ...**, **THE SAME NUMBER OF ... AS ...**

Standard Quantifiers: Some Linguistic Generalizations

We consider cases where the quantifier semantics sketched above has proven enlightening in formulating linguistic generalizations. As a first case, observe that in (4a) the presence of *ever*, called a ‘negative polarity item (npi),’ is ungrammatical; in (4b) it is fine.

- (4a) *Some student here has ever been to Pinsk.
 (4b) No student here has ever been to Pinsk.

Similarly *any* in *No child saw any birds on the walk* is an *npi*, not licensed in **Some child saw any birds on the walk*. **The linguistic query:** Which NPs in contexts like (4) license *npi*'s? To within a good first approximation, the answer is given by the Ladusaw-Fauconnier Generalization (Ladusaw, 1983): *The NPs which license npi's are just those which denote decreasing (order reversing) generalized quantifiers*. A generalized quantifier *F* is *decreasing* iff whenever $A \subseteq B$ then if $F(B) = T$ then $F(A) = T$. **No student** is decreasing since if all *As* are *Bs* ($A \subseteq B$), then if no student is a *B*, it follows that no student is an *A* (otherwise that student would also be a *B*, contrary to assumption). In contrast, **some student** is not decreasing: perhaps all *As* are *Bs* and there are many students among the *Bs* but they all lie among those that are not *As*. The reader can verify that the following NPs are decreasing and do license *npi*'s: *fewer than five students, less than half the students, not more than two students, neither John nor Bill, no student's doctor*. In contrast, NPs such as *John, more than five students, most poets, either John or Bill, some student's doctor* are not decreasing and do not license *npi*'s. Keenan and Szabolcsi (see **Boole and Algebraic Semantics**) point out that standard negation is also a decreasing function, and it also licenses *npi*'s: **John has ever been to Pinsk* vs. *John hasn't ever been to Pinsk*.

Our first generalization also illustrates that whether an NP of the form [Det + N] is decreasing or not is decided by the Det. So if *no student* is decreasing then, so is *no child, no professional acrobat*, etc. since they all have the same Det. Similarly if *some student* is not decreasing, then neither is *some child, some professional acrobat*, etc.. Many linguists refer to the expressions we call NPs as DPs ('Determiner Phrases'), in part because significant properties of the entire expression, such as whether it is decreasing or not, is decided by the choice of Det. (But other properties, such as whether the expression is animate, feminine, or satisfies the selection restrictions of a predicate, are determined by the N. #*Every idea laughed* is bizarre, since ideas aren't the kinds of things that can laugh and changing the Det does not improve matters: #*Some idea laughed, #Most ideas laugh*).

As a second generalization, consider how we may characterize the NPs that are *definite* (plural) in the sense that they may grammatically replace *the poems* in *two of the poems*. Some such NPs are *the ten poems, these (ten) cats, John's (ten) students*. Some NPs that are not definite (plural) in this sense are: *no poems, every cat, most students*. Again, to within a

first approximation, we may say that the definite plural NPs are those of the form [Det + N], where (in each domain) Det denotes a function *f* that satisfies:

- (5) For all *A*, either $f(A)(B) = F$; all *B*, or for some $X \subseteq A$, $f(A)(B) = T$ iff $X \subseteq B$.

To illustrate the idea, **(John's ten) (students)** maps each set *B* to *F* if John does not have exactly ten students; if he does, then it maps a set *B*, such as **daydream in class** to *T* iff the property of being a student-which-John-has is a subset of *B*. See Barwise and Cooper (1981) and Matthewson (2001) for further discussion.

The two generalizations adduced so far are semantic characterizations of syntactic phenomena. For purposes of defining the class of well-formed expressions in English, we need to know which NPs license *npi*'s in the predicate and which may occur naturally in the post of position in partitives. Most approaches to generative grammar desire a purely syntactic definition of these classes. Our observations do not rule out such definitions. Indeed, they provide a criterion for whether a proposed definition is adequate or not. Still, at the time of writing, we have no explicit syntactic definition of these two classes of NPs.

A third problem comparable to our first two is to characterize those NPs that occur naturally in *Existential There* (ET) contexts, as in (6).

- (6a) There are/aren't *more than ten boys* in the room.
 (6b) *There are/aren't *most boys* in the room.

This problem has vexed generative grammarians since Milsark (1977). See Reuland and ter Meulen (1987) and Keenan (2003). Part of the problem is that affirmative declarative Ss of the form in (6) have a variety of uses, with different uses seemingly allowing different NPs. For example, in so-called 'list contexts' (Rando and Napoli, 1978), ET sentences admit definite NPs such as *the bus* in *How can I get to UCLA from here? Well, there's always the bus, but it doesn't run very often*. But mostly these uses are not preserved in negative or interrogative Ss. Putting aside uses limited to affirmative declarative Ss, Keenan (2003) supports that NPs that occur freely in ET contexts are ones built from Dets that denote a certain kind of 'conservative' function. A Det is, standardly, **conservative**, or as we shall say here, **conservative on its first argument**, if it satisfies (7a), stated more generally in (7b).

- (7a) Det poets daydream iff Det poets are poets who daydream
 (7b) $D(A)(B) = D(A)(A \cap B)$

To see that *most but not all* is conservative, for example, check that *Most but not all poets daydream*

and *Most but not all poets are poets who daydream* always have the same truth value. Indeed, the second sentence seems redundant, with the predicate just repeating information already contained in the Noun argument.

The formulation in (7b) says that in evaluating the truth of $D(A)(B)$, we may limit the predicate argument B to those of its elements that occur in A . In this sense, then, we shall say that D is CONS1, ‘conservative on its first argument.’ And we support:

(8) Cons in general Dets denote CONS1 functions

Cons is a new generalization, not one that arises directly in response to a query from independent linguistic study. It is surprisingly strong. Given a domain E with n elements, Keenan and Stavi (1986) show that for $k = 4^n$ there are 2^k functions from pairs of properties to $\{T, F\}$. Only $2^{k'}$ of these functions are Cons, where $k' = 3^n$. So in a domain with just two individuals, there are $2^{16} = 65,536$ maps from pairs of properties to truth values, only $2^9 = 512$ of which are conservative! So Cons rules out most logically possible denotations for Dets. Here is a simple non-conservative function, F : $F(A)(B) = T$ iff $|A| = |B|$. Clearly $F(\{a, b\}, \{a, c\}) = T$, but $F(\{a, b\}, \{a, b\} \cap \{a, c\}) = F(\{a, b\}, \{a\}) = F$, so F fails Cons.

Now, mathematically it makes sense to ask whether there are Dets that are conservative on their second, predicate, argument. Such Dets would denote functions D satisfying $D(A)(B) = D(A \cap B)(B)$, where we can limit the A s we consider to those that lie in B . Many natural classes of Dets fail CONS2. For example, **universal** Dets, such as *all*, *all but ten*, *every...but John*. It might be false that all poets daydream, but it must be true that all poets who daydream daydream. So *all* is not CONS2. Also not CONS2 are **definite** Dets, such as *the*, *the ten*, *these (ten)*, *John’s (ten)*; and **proportional** Dets, such as *most*, *half (of) the*, and *not one...in ten*. But there is one large class of Dets that are CONS2. They include the *intersective* Dets, ones whose values just depend on which objects have both the Noun property and the Predicate property. For example, *some* is intersective since the truth of $SOME(A)(B)$ is decided just by checking $A \cap B$ (verifying that it is not empty). Let us define:

(9) D is *intersective* iff for all A, A', B, B' if $A \cap B = A' \cap B'$ then $D(A)(B) = D(A')(B')$

So intersective D s are ones ‘invariant’ under replacement of A and B with other arguments A' and B' provided the intersection of the pairs of arguments remains unchanged. Note that, an intersective D is necessarily CONS 1: since $A \cap B = A \cap (A \cap B)$ we infer $D(A)(B) = D(A)(A \cap B)$. And since $A \cap B =$

$(A \cap B) \cap B$, we have that $D(A)(B) = D(A \cap B)(B)$, and so intersective Dets are CONS2. In fact:

Theorem D is intersective iff D is CONS1 and CONS2 (Keenan, 2003)

An important special case of intersective Dets are **cardinal** ones, whose values depend on the cardinality of $A \cap B$, such as *at least n* , *more/fewer than n* , *at most n* , *approximately n* , *between n and m* , *several*, *a dozen*, *just finitely many*, and *infinitely many*. And we now answer our third query (Keenan, 2003):

(10) NPs which occur freely in Existential There contexts are (boolean compounds of) ones built from CONS2 Dets.

‘Boolean compounds’ here just refers to NPs built by conjunction, disjunction, and negation. For example, *There are at least two dogs and not more than five cats in the garden* is fine since *at least two dogs* and *more than five cats* both occur in ET contexts.

Example (10) predicts that NPs built from intersective Dets occur in ET contexts, and this is correct: *Aren’t there between five and ten students in your class?* *Was there no student (but John) in the building?* etc. (Note that **no...but John** treated as a Det is intersective, as *No A but John is a B* is True iff $A \cap B = \{\text{John}\}$, that is, the only A that is a B is John). Similarly, the universal, definite, and proportional Dets noted earlier are predicted not to occur freely in ET contexts since they fail CONS2, and this is correct: **Weren’t there all/most/the students in the class?*

Are there are CONS2 Dets that fail to be intersective? By the theorem, they would have to fail to be CONS1, and so rather rare. But there are two candidates: NPs of the form [only/mostly N], such as *only poets*. Interpreting *only* as a Det yields $ONLY(A)(B) = T$ iff $B \subseteq A$. For example, *Only poets daydream* is true iff everyone who daydreams is a poet (but there may be poets who don’t daydream). Clearly $ONLY$ thus defined is CONS2, since $ONLY(A \cap B)(B) = T$ iff $B \subseteq A \cap B$, iff $B \subseteq A$. So (10) predicts the well-formedness of *There weren’t only poets at the party*, which is correct.

A last case covered, unexpectedly, by (10) are NPs built from comparative Dets like *more...than...* which combined with two N s to form an NP. Each N property is a conservativity domain. To evaluate whether *More students than teachers daydream* we must consider both the students who daydream and the teachers who daydream. But not only are comparative NPs conservative on their two N arguments; they are intersective, in fact, cardinal. To decide whether more A s than B s have C , we need only check $|A \cap C|$ and $|B \cap C|$, verifying that the former is greater than the latter. Thus, (10) predicts

that cardinal comparatives should occur in ET contexts, and they do: *Weren't there more students than teachers at the party?* (We can also compare predicate properties, as in *More poets drink than smoke*, in which case only the single N property is a conservativity domain).

A fourth linguistic generalization is given in (11), where 'lexical' means 'not syntactically derived.'

- (11) Lexical NPs are monotonic, and lexical Dets build monotonic NPs.

A monotonic NP is one that either denotes a decreasing function (already defined) or an increasing one, where a generalized quantifier *F* is **increasing** iff whenever $A \subseteq B$ then if $F(A) = T$ then $F(B) = T$. Lexical NPs, principally proper nouns (*John, ...*) and pronouns (*he, she, ...*), are easily seen to be increasing: to paraphrase Aristotle, if all poets daydream and Paul is a poet then Paul daydreams. The reader may verify that *all/some/most/the five/my/John's cats* are all increasing, so lexical Dets usually build increasing NPs. But *no* and *neither* build decreasing ones. The only slightly doubtful case are NPs built from bare numerals, such as *two poets*. The NP is increasing if *two* is understood in the sense of *at least two*; it is not monotonic if understood as *exactly two*. We find the *at least two* reading the most natural because in some cases that is clearly what is intended, as in *Are there two free seats in the front row?* Additional information from context can be invoked to force the *exactly two* reading. Our last generalization in this section is (12):

- (12) Natural Language Dets are Domain Independent

In practical applications, such as in database theory (Abiteboul *et al.*, 1995), it is important to know what the domain of objects is that properties under consideration are subsets of. It would be more accurate (van Benthem, 1984) to treat NPs, for example, as functions that associate with each possible domain *E* a generalized quantifier over *E*, a function associating with each subset *A* of *E* a truth value. Similarly Dets associate with each *E* a function mapping subsets of *E* to generalized quantifiers over *E*. But (12) says that the value a Det denotation *D* assigns to a pair *A, B* of properties cannot depend on the choice of underlying domain (as long as *A* and *B* are subsets of that domain). This (seemingly vacuous) constraint entails, for example, that natural languages could not present a Det *blik* defined by: *Blik As are Bs* is True iff the number of non-*As* is two. The truth value of such an *S* would vary with the domain: if $A = \{a\}$ and $E = \{a, b, c\}$ then it is True, but if $E = \{a, b\}$ it is False.

To close this section, we note that our treatment of NPs and Dets enables us to specify precisely certain traditional, if informally given, classes of expressions. We note two cases. First, the linguistic literature on quantification usually builds heavily on the specific quantifiers *some* and *all*, which we have already defined. But now we can see that they represent two quite general classes of Dets: the *Existential* ones, including *some*, are just those that denote intersective functions, as defined in (9). The *Universal* Dets are those that denote cointersective ones, as defined in (13).

- (13) *D* is *co-intersective* iff for all *A, B, X, Y* if $A - B = X - Y$ then $D(A)(B) = D(X)(Y)$

Thus, whether a cointersective function *D* is True of a pair *A, B* is decided just by checking $A - B$. To see that ALL is cointersective observe that, as defined, $ALL(A)(B) = T$ iff $|A - B| = 0$. (*ALL BUT TEN*) (*A*) (*B*) = *T* iff $|A - B| = 10$, etc. Our approach also enables us to see that proportionality Dets such as *most, half, more than one ... in ten*, etc., are more complicated than either intersective or cointersective Dets, because their truth at a pair *A, B* of properties depends on both $A \cap B$ and $A - B$.

Secondly, most of the Dets discussed in the linguistic literature are ones with a 'logical' or 'mathematical' sense: *some, every, most, most of the ten, not all, most but not all, between a third and two thirds of the*, etc. But we have also countenanced a few which are more 'empirical', such as *my, John's ten, no ... but John*, etc. (see Keenan, 1996). Can we say in precise terms what the distinction is? We can. The 'logical' Dets are those that are invariant under permutations of the elements of the underlying domain *E*. A permutation of *E* is simply a one-to-one function from *E* onto *E*. If *h* is such a function and *A* a subset of *E*, then by $h(A)$ is meant $\{h(x) | x \text{ in } A\}$. And we say that a generalized quantifier *F* is *permutation invariant* (PI) iff for all *A*, all permutations *h* of *E*, $F(A) = F(h(A))$. A Det function *D* is PI iff for all *A, B* and all permutations *h*, $D(A)(B) = D(h(A))(h(B))$. Note that, for any $X \subseteq E$, *X* and $h(X)$ always have the same cardinality. So the 'logical' quantifiers are those that cannot distinguish between properties *A, A'* of the same cardinality (and whose complements, $E - A$ and $E - A'$ have the same cardinality when *E* is infinite). One computes, then, that Dets such as *some, every, most, most of the ten, not all, most but not all, between a third and two thirds of the* always denote PI functions, whereas ones like *my, John's ten*, etc., do not.

Some Non-Standard Quantifiers

We begin by considering the interpretation of the quantified NPs already discussed when they occur as

objects of transitive and ditransitive verbs: *John envies all movie actors*, *He gave most of his teachers several presents*. Many semanticists regard these uses as illustrating a ‘type mismatch.’ To interpret the object NP in *John interviewed every applicant* we apparently need to treat it as a function mapping a binary relation (**interview**) to a property, but we are already committed to interpreting NPs as maps from properties (unary relations) to truth values (zero-ary relations). But in fact there is no problem here at all. We know exactly what property *interviewed every applicant* denotes. It is the set of objects x such that ‘ x interviewed every applicant’, that is, such that **every applicant** is true of the *set* of objects that x interviewed. In the last phrase, we are applying the generalized quantifier **every applicant** to a set – which is exactly how we have already been interpreting NPs. So the value that an NP denotation assigns to a binary (ternary, ...) relation is determined by the value that it assigns to sets (properties). So the solution to the ‘type mismatch’ problem is to treat NPs as directly denoting functions mapping $n+1$ -ary relations to n -ary ones in such a way that their values at $n > 1$ -ary relations are determined by their values at the unary relations. Here is the solution for binary relations. Keenan and Westerståhl (1997) give the general statement. We write aR for $\{b | aRb\}$, the set of objects that a stands in the relation R to.

- (14) A type $\langle 1 \rangle$ function F over a universe E maps each subset of E to a truth value and each binary relation R over E to a subset of E by:
 $F(R) = \{a | F(aR) = T\}$.

Clearly, then, each generalized quantifier uniquely determines a type $\langle 1 \rangle$ function; all we have done is add more objects (binary relations) to its domain. So the interpretation of an S with two quantified NPs is given compositionally, as in (15).

- (15a) No politician kissed every baby
 (15b) (no politician) ((every baby) (kissed))

Note that, (15b) means ‘No politician has the property that he kissed every baby,’ the object narrow scope (ONS) reading of (15a). In fact, (15a) does not have an object wide scope (OWS) reading, on which it would mean that every baby has the property that no politician kissed him. But some S s with two quantified NPs do present such ambiguities, a matter of much concern to linguists (Szabolcsi, 1997). *Some editor read every manuscript* has an ONS reading, representable analogously to (15b), on which it means that there is an editor with the property that he read every ms. But it also has a OWS reading, on which it means that every ms has the property that some editor read it – so the editors may vary with the manuscripts. An easy use of

variable binding operators (VBOs) allows us to represent the less accessible OWS reading by:

- (16) (every manuscript x) ((some editor) (read x))

The use of VBOs has distracted us from the fact that the range of logically possible interpretations of nonsubject NPs is vastly greater than of subject NPs. Given a universe E with more than one element, there are many more functions from binary relations to properties than there are from properties to truth values. And natural languages provide the means for denoting some of these. To see this, we need a way to test functions H from relations to properties to see if they are possibly extensions of functions from properties to truth values. Here is such a test, given first by example, where we test X :

- (17a) If John praised exactly the people who Bill criticized then John praised X iff Bill criticized X
 (17b) If $aR = bS$ then $a \in F(R)$ iff $b \in F(S)$

For example, *most of the Peter’s students* passes the X test in (17a). Given the truth of the *if*-clause, we infer that *John praised most of Peter’s students* and *Bill criticized most of Peter’s students* have the same truth value. (This is hardly surprising; the NP we are testing occurs as a subject of a P1 and thus is interpretable as a generalized quantifier.)

But consider now the reflexive pronoun *himself*. It fails the test. Imagine, for example, that the *if*-clause is true and that John praised just Sam, Frank, Bill, and Sue. Then those are just the people that Bill criticized, so *Bill criticized himself* is true, but *John criticized himself* is false. Hence, there is *no* function from properties to truth value that takes exactly the value on binary relations as *himself* does. That is, reflexives represent an increase in logical expressive power. Moreover *himself* is not unique here; all (nontrivial) referentially dependent NPs in object position are logically new in this sense. This includes ones like *everyone but himself*, *both himself and the teacher*, *everyone smarter than himself* (as in *John criticizes everyone smarter than himself*) in which the NP must be referentially dependent, as well as NPs like *his mother* where it simply may be dependent. It is the dependent interpretation where it is new.

The increase in logical expressive power afforded by nonsubject NPs is far more extensive than instantiated by the referentially dependent NPs. They are not functions of type $\langle 1 \rangle$, but they can be correctly interpreted as functions from binary relations to sets, (just not ones that extend appropriately a generalized quantifier). But the sort of dependency in (18) cannot be handled in a comparable way:

- (18) Different people like different things

The weakest truth conditions of (18) are easy to state: for any two (different) people x and y , the set of things that x likes is not identical to the set that y likes. And we know (Keenan, 1996) that over any domain with several people and things, there are *no* type $\langle 1 \rangle$ functions F and G such that for all binary relations R , $F(G(R))$ is true iff different people stand in the relation R to different things. Similar claims hold for other S s which involve comparing different object sets with different choices of subject argument: *John and Bill support rival political parties*, *Rosa and Zelda date men who dislike each other*, etc. In fact, the same claim holds even when the comparison is not one of difference: *All the students answered the same questions on the exam*, *They wore the same color necktie*, etc. Speaking very informally, we can say that the combination of subject and object expressions in these S s place conditions on the relation denoted by the transitive verb which are inherently ‘relational,’ not expressible as independently statable conditions on each argument. To state this more explicitly, let us define:

- (19) A function H from binary relations to truth values is said to be of type $\langle 2 \rangle$. Such a function is type $\langle 1 \rangle$ *reducible* iff there are type $\langle 1 \rangle$ functions F, G such that for all binary relations R , $H(R) = F(G(R))$.

Then what we are saying above is that the type $\langle 2 \rangle$ functions expressed by (Different people, different things), (John and Bill, rival political parties), etc., are not type $\langle 1 \rangle$ reducible. Thus, they are not expressible as the composition of two generalized quantifiers, and so are logically new. Some additional examples are induced by the *else-else* construction in (20), the *which-which* construction in (21), reciprocal objects in (22), and predicate anaphors in (23). For further examples, see Keenan (1996).

- (20) John didn’t praise Bill but everyone else praised everyone else.
John praised Bill but no one else praised anyone else.
- (21) John doesn’t know which students answered which questions on the exam.
- (22) The students were shouting at each other.
- (23) John read more books than Bill (did).

In sum, we have a list of expression types that are not type $\langle 1 \rangle$ reducible, but we do not know precisely what type $\langle 2 \rangle$ functions are expressible. All of them (over finite E)?

Thus, much remains to be discovered even in English, the language in which quantifiers have been the most extensively studied. And we are really just beginning to study quantification in less well known

languages (Bach *et al.*, 1995; Matthewson, 2001) as well as in contexts other than that of NPs and Dets, the most prominent here being temporal and event quantification, using adverbial quantifiers, as in *Matt always/often/occasionally/seldom/rarely/never visits museums on weekends* (de Swart, 1996).

Answers to Exercises

(AT LEAST TWO) (A) (B) = T iff $|A \cap B| \geq 2$

BOTH(A) (B) = T iff $|A| = 2$ and $A \subseteq B$

(AT MOST FIVE OF THE TEN) (A) (B) = T iff

$|A| = 10$ and $|A \cap B| \leq 5$

(LESS THAN HALF THE) (A) (B) = T iff

$|A \cap B| < |A|/2$

See also: Boole and Algebraic Semantics; Formal Semantics; Monotonicity and Generalized Quantifiers.

Bibliography

- Abiteboul S, Hull R & Vianu V (1995). *Foundations of databases*. Reading, MA: Addison-Wesley.
- Bach E, Jelinek E, Kratzer A & Partee B H (eds.) (1995). *Quantification in natural languages*. Dordrecht: Kluwer.
- Barwise J & Cooper R (1981). ‘Generalized quantifiers and natural language.’ *Linguistics and Philosophy* 4, 159–219.
- Beghelli F (1993). ‘Structured quantifiers.’ In Kanazawa & Piñón (eds.), 119–147.
- de Swart H (1996). ‘Quantification over time.’ In van der Does & van Eijck (eds.), 311–337.
- Gärdenfors P (ed.) (1987). *Generalized quantifiers*. Dordrecht: Reidel.
- Kanazawa M & Piñón C J (eds.) (1994). *Dynamics, polarity, and quantification*. Stanford: CSLI.
- Keenan E L (1996). ‘Further beyond the Frege boundary.’ In van der Does & van Eijck (eds.), 179–203.
- Keenan E L (1996). ‘The semantics of determiners.’ In Lappin S (ed.) *The handbook of contemporary semantic theory*. Cambridge, MA: Blackwell. 41–63.
- Keenan E L (2003). ‘The definiteness effect: semantics or pragmatics?’ *Natural Language Semantics* 11(2), 187–216.
- Keenan E L & Moss L M (1985). ‘Generalized quantifiers and the expressive power of natural language.’ In van Benthem & ter Meulen (eds.), 73–127.
- Keenan E L & Stavi J (1986). ‘A semantic characterization of natural language determiners.’ *Linguistics and Philosophy* 9, 253–326.
- Keenan E L & Westerståhl D (1997). ‘Generalized quantifiers in linguistics and logic.’ In van Benthem J & ter Meulen A (eds.) *Handbook of language and logic*. Amsterdam: Elsevier. 837–895.
- Krynicky M, Mostowski M & Szczerba L W (1995). *Quantifiers, logics, models and computation* (vol. 1). Dordrecht: Kluwer.
- Ladusaw W (1983). ‘Logical form and conditions on grammaticality.’ *Linguistics and Philosophy* 6, 389–422.

-
- Lappin S (1988). 'The Syntax and semantics of NPs.' *Linguistics* (special issue) 26.6. 903–909.
- Matthewson L (2001). 'Quantification and the nature of crosslinguistic variation.' *Natural Language Semantics* 9, 145–189.
- Milsark G (1977). 'Toward an explanation of certain peculiarities of the existential construction in English.' *Linguistic Analysis* 3, 1–29.
- Rando E & Napoli D J (1978). 'Definites in THERE-sentences.' *Language* 54, 300–313.
- Reuland E & ter Meulen A (1987). *The representation of (in)definiteness*. Cambridge, MA: MIT Press.
- Szabolcsi A (1997). *Ways of scope taking*. Dordrecht: Kluwer.
- van Benthem J (1984). 'Questions about quantifiers.' *Journal of Symbolic Logic* 49, 443–466.
- van Benthem J & ter Meulen A (eds.) (1985). *Generalized quantifiers in natural language*. Dordrecht: Foris.
- van der Does J & van Eijck J (eds.) (1996). *Quantifiers in natural language*. Stanford: CSLI.
- Westerståhl D (1989). 'Quantifiers in formal and natural languages.' In Gabbay D & Guenther F (eds.) *Handbook of philosophical logic*, vol. 4. Dordrecht: Reidel. 1–133.
- Westerståhl D (1995). 'Quantifiers in natural language: a survey of some recent work.' In Krynicki *et al.* (eds.). Dordrecht: Kluwer. 359–408.
-

This page intentionally left blank

R

Radical Interpretation, Translation and Interpretationalism

M Hymers, Dalhousie University, Halifax, Nova Scotia, Canada

© 2006 Elsevier Ltd. All rights reserved.

Donald Davidson's idea of **radical interpretation** gets its inspiration from W. V. Quine's account of **radical translation**. This article will deal first with Quine's view and then Davidson's before turning to the **interpretationalism** of Davidson's views and criticisms of thereof.

Radical Translation

W. V. Quine is generally suspicious of such mental notions as belief, intention, and meaning, preferring an austere behaviorist account of language use. Quine imagines a field linguist attempting to compile a **translation manual** that would pair sentences of her own language with sentences of a language wholly unlike any she has ever encountered, unaided by dictionaries or local interpreters. In this scenario of **radical translation**, the linguist has access only to the bodily behavior of the speaker and to those objects and events that are manifest in the local environment. The best pragmatic procedure for the radical translator to follow, thinks Quine, is to try to correlate the speaker's utterances with their proximal stimuli – “surface irritations” (Quine, 1960: 22) of sensory receptors. If a given stimulus—say, that produced by a rabbit—can be correlated with a given utterance on one occasion – say, ‘Gavagai!’ – then the translator must try to reproduce similar stimuli in order to see whether they provoke similar utterances (find some more rabbits and point them out). Or, better, the translator may on various stimulus-occasions, similar and different, repeat the speaker's original utterance in the hope of eliciting terms for assent and dissent (point to rabbits and rutabagas alike, asking, ‘Gavagai?’). With a firm working hypothesis about these terms in hand, she may proceed to pair sentences of the speaker's language with sentences of her own language, according as they seem from her perspective to have the same **stimulus meanings** – that is, roughly, to be

prompted by the same sorts of events and objects. Hypotheses concerning assent and dissent may, of course, meet with disconfirming evidence, forcing the translator to begin anew with another hypothesis. But by following this procedure, Quine thinks, the field linguist can expect eventually to translate **observation-statements** – reports on observable occurrent events or states of affairs. By applying a **principle of charity**, which counsels the translator not to assume wanton irrationality on the speaker's part, she may also hope to translate logical terms, such as ‘not,’ ‘and,’ and ‘or,’ since a rational speaker is unlikely to affirm each conjunct of a conjunction (‘Roses are red’ and ‘Violets are blue’) without also affirming the conjunction (‘Roses are red, **and** violets are blue’). Efforts to take radical translation further to relatively nonobservational sentences – such as ‘Electrons have negative charge’ or ‘Incest is an abomination’ – will rely on the formulation of **analytical hypotheses** regarding how to parse the speaker's utterances into words and other sub-sentential components. However, Quine contends, there will always be rival systems of analytical hypotheses, any of which will be as pragmatically successful as the other for purposes of compiling a workable translation manual. Translation is thus indeterminate, but Quine thinks that the lessons of radical translation apply with equal justice to our attempts to converse with our colinguists and to our learning of our own language. Having learned the language already, we find the procedure of interpreting our fellow speakers and ourselves as familiar as can be, but in neither case do we have any more evidence to go on than is available to the field linguist. It follows, thinks Quine, that meaning itself is no more determinate than translation (*see Indeterminacy, Semantic*).

Radical Interpretation

In the work of Donald Davidson, radical translation is transmuted into **radical interpretation**. The two notions differ in at least three important respects. (1) Whereas the former notion rests on methodological behaviorism, the latter does not. (2) Whereas radical translation aims to compile a translation

manual for the language under study, radical interpretation aims to formulate and test, as part of a general theory of interpretation for a speaker, a core meaning theory for the speaker's language, which will describe the finite vocabulary of the language and show how that vocabulary can be combined using a finite set of recursive rules to produce any sentence in the language. Knowledge of such a theory would supposedly "suffice" (Davidson, 1984: 172; 1986: 438) for understanding the speaker, but is not necessary (1984: 125; 1986: 438). (3) Whereas radical translation relies on the principle of charity methodologically to argue for the translatability of logical connectives, radical interpretation applies this principle generally as something like a necessary condition of the very possibility of interpretation (Ramberg, 1989: 77). These points need elaboration.

(1) Quine focuses on terms for assent or dissent "because they suggest a behaviouristic test," but Davidson has no such commitment to behaviorism (Davidson, 1984: 231). He criticizes Quine's preoccupation with proximal causes of sensory stimulation, arguing that Quine slips back into a Cartesian model of the mind as isolated from the world, thereby exposing himself to both an "old-fashioned skepticism of the senses" (Davidson, 1990: 74) and a relativism about truth that arises from the fact that no two individuals share the same patterns of sensory stimulation (1990: 76). Translation even of observation sentences, which Davidson does not distinguish sharply from theoretical sentences (2001: 149) can be assured to work only if the interpreter focuses on macroscopic objects and events in a world shared by interpreter and speaker – marmots and mangoes, instead of photons striking nerve endings. In order to locate the source of the speaker's sensory stimulations, we need, says Davidson, three patterns of similarity from one occasion to another: the similarity I find in my experiences, the similarity the speaker finds in her experiences, and the similarity that I find in the speaker's responses to her experiences (2001: 119). The interpreter must "correlate his own responses and those of the speaker by reference to the mutually salient causes in the world of which they speak" (1990: 78) – a process Davidson analogizes to 'triangulation' (2001: 105). (See 'Interpretationalism' below for more on this process.)

(2) The notion of translation does not by itself, thinks Davidson, give a plausible account of meaning for the speaker's language. A better account would have to show systematically how "the meanings of sentences depend upon the meanings of words" (2001: 17) if a speaker's capacity to understand and produce a potential infinity of novel and banal utterances is to be represented (*see Compositionality:*

Philosophical Aspects). Drawing on the formal logical work of Alfred Tarski (1949), Davidson proposes that we begin the task of interpretation by taking as basic the attitude of holding a sentence true (or false). Then, having identified the sentences a speaker holds true, we must formulate hypotheses about what finite vocabulary and what rules of syntax are operating in the speaker's language, and what rules of satisfaction are descriptive of the various predicates in the language. (A rule of satisfaction tells us what things satisfy a given predicate – roughly, of what things that predicate is true. It tells us which are the blue things or the koalas, for example.) Such a procedure provides a "model" (Davidson, 1986: 438) of what we commonly do when interpreting a speaker, though we cannot be plausibly said to **know** any such theory of interpretation (1986: 438). What exactly this means has occasioned some puzzlement (see Dummett, 1975, 1976; Schiffer, 1989: 116–117).

(3) This procedure requires that we be able to identify which sentences a speaker holds true, but this by itself will not tell us what those sentences mean. If we knew what the speaker believed in the circumstances, then that would give us some substantial clues concerning what she means by her sentences. However, our surest way of identifying what a speaker believes is by first knowing what her words mean. To escape this endless circle, Davidson appeals to the principle of charity, which, in his formulation, has it that we should "assume general agreement on beliefs" (1984: 196) and "try for a theory that finds [the speaker] consistent, a believer of truths, and a lover of the good (all by our own lights, it goes without saying)" (1980: 222). This is not a principle that we choose simply in order to avoid making our speaker sound "silly" (Quine, 1960: 59). Rather, "charity is forced on us" (Davidson, 1984: 197) if we are to be in a position to interpret our speaker **at all**.

Davidson believes that meaning- and attitude-attribution are holistic by nature. There is no sense to the idea of attributing a belief to a creature unless one attributes "a world of further beliefs to give it content and identity" (Davidson, 2001: 98) and likewise no way of assigning meaning to a sentence in isolation from assignments of meaning to other sentences (1980: 239). Such holism entails the possibility of reassigning meaning and content, so long as one makes compensating adjustments elsewhere in the attributed set of background meanings and beliefs. Thus Davidson allows for an indeterminacy of interpretation that is the "semantic analogue of Quine's indeterminacy of translation" (1984: 225), but as a consequence of his broader application of the principle of charity, he believes that "the range of indeterminacy is less than Quine thinks it is" (1984: 228).

What indeterminacy remains is, in Davidson's view, comparable to the possibility of adopting different scales of measurement for the same objective phenomenon, much as we may measure temperature on the Fahrenheit or Celsius scales (1984: 154, 225; 1980: 257; 2001: 145). Additionally, Davidson regards it as a facet of linguistic competence that one has "first person authority" (2001: 3) concerning the meanings of one's own words and the contents of one's own thoughts. So, unlike Quine, he does not think radical interpretation "begins at home" (Quine, 1969: 46) (though it might begin next door when one interprets one's colinguists).

Interpretationalism

Davidson's account of radical interpretation is readily coupled to a view about meaning and intentional content known as **interpretationalism**. Much as Quine maintains that translation is prior to meaning and not vice versa, as we might have thought prereflectively, Davidson contends that there is no fact concerning what a speaker's words mean or what the contents of her intentional attitudes are apart from the possibility of their meaning or content being attributed by an interpreter. This is because, as we saw briefly above, interpreting a speaker requires identifying the mutually salient causes of her verbal behavior. Any link in the causal chain extending from a publicly observable object to the surface irritations of her sensory organs could, on some account, be identified as the cause if the speaker is left to try to interpret herself. From a third-person point of view, however, what seems explanatory of the speaker's behavior is her pattern of responses to public objects, to which we also respond. "It takes two points of view to give a location to the cause of a thought, and thus to define its content" (Davidson, 2001: 212–213).

Criticisms of Radical Interpretation and Interpretationalism

Critics of interpretationalism may accuse it of an implausible instrumentalism about the mental (Davidson, 2001: 70; Hookway, 1988: 181), according to which intentional attitudes are no more than convenient fictions that we employ when trying to predict and explain the behavior of a rational creature. Furthermore, interpretationalism seems to lead to a vicious regress, since one's possession of meanings and intentional attitudes requires being interpreted by another, whose intentional attitudes, in turn, must be derived from the interpretations of a third, and so on (Heil, 1998: 131). Similar criticisms

are made of Daniel Dennett's "intentional stance" (1971; 1987) – see Putnam, 1987: 15–16; Rudder-Baker, 1989: 305 – a variety of "selective instrumentalism" (Dennett, 1987: 72), which Dennett distinguishes from "interpretationism" (1987: 15) and "classical instrumentalism" (1987: 71). Davidson (2001: 82) thinks that Dennett confuses the question of whether intentional attitudes are entities with the question of whether there are correct answers about what intentional attitudes anyone has. But Davidson replies here that his account is no more instrumentalist than are systems of measurement (2001: 74–75). The fact that we can measure temperature differently in Fahrenheit or Celsius does not show that no objective phenomenon is being measured; Dennett makes a similar response to the parallel criticism of his view (1987: 34). And the apparent regress arises only if we fail to see practices of interpretation as coming into being gradually and socially, producing interpretative communities "once they reach an appropriate level of sophistication" (Heil, 1998: 153).

Critics have doubted Davidson's ability to account for the meanings of sentences that, intuitively, are neither true nor false, such as interrogatives and imperatives, and, perhaps more seriously, of such sentences as direct and indirect quotations and attributions of intentional attitudes. (Tarski himself thought that his treatment of truth, from which Davidson borrows, could not be applied to natural languages.) Davidson has long recognized that there are programmatic challenges of this sort for his view to meet (1984: 35–36) and has taken an optimistic view of his chances, viewing the problems as a series of puzzles to be dealt with by an active research program (see the various essays in Davidson, 1984), though many critics remain pessimistic (Glock, 2003: 159–165; Schiffer, 1989).

Davidson's use of the principle of charity has also been criticized on a number of grounds. Some think it smacks of cultural imperialism, forcing us to construe cultural others as like their interpreters in all essential respects (MacIntyre, 1985). Davidson may respond that meaningful disagreement requires a broad background of agreement and that cultural difference cannot even be recognized without starting with something like the principle of charity (1984: 196–197; 2001: 39–41, 148–149). Some critics have suggested that charity is not sufficient for interpretation, since we have reason not to attribute what we take to be true beliefs to a speaker if it would be mysterious how she came to hold such a belief (Grandy, 1973: 445), and have proposed in its place a "principle of humanity" (Grandy, 1973: 443) according to which we should attribute beliefs that we would find it

reasonable to hold in the speaker's position. Davidson has responded with attempts to clarify the principle of charity (1984: xvii, 136) and with a proposal to distinguish two principles, Coherence and Correspondence (2001: 211). Others contend that it is shared behavior, not merely agreement in beliefs, that makes interpretation possible and that successful interpretation can thus succeed in the absence of charity (Glock, 2003: 194–199), but although such behavioral commonality may not be premised in the principle of charity, Davidson seems to agree that it is of central importance (2001: 128–129, 212–213), and the difference here is **perhaps** one of emphasis.

See also: Compositionality: Philosophical Aspects; Indeterminacy, Semantic; Paradoxes, Semantic; Truth: Theories of in Philosophy.

Bibliography

- Burwood S, Gilbert P & Lennon K (1999). *Philosophy of mind*. Montreal: McGill-Queen's University Press. Chapter 5.
- Davidson D (1980). *Essays on actions and events*. Oxford: Clarendon Press.
- Davidson D (1984). *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
- Davidson D (1986). 'A nice derangement of epitaphs.' In LePore E (ed.) *Truth and interpretation: perspectives on the philosophy of Donald Davidson*. Oxford: Basil Blackwell. 433–446.
- Davidson D (1990). 'Meaning, truth and evidence.' In Barrett R & Gibson R F (eds.) *Perspectives on Quine*. Oxford: Blackwell. 68–79.
- Davidson D (2001). *Subjective, intersubjective, objective*. Oxford: Clarendon Press.
- Dennett D (1971). 'Intentional systems.' *The Journal of Philosophy* 68(4), 87–106.
- Dennett D (1987). 'True believers: the intentional strategy and why it works.' In *The intentional stance*. Cambridge, MA: MIT Press. 13–35.
- Dummett M (1975). 'What is a theory of meaning? (I)' In Guttman S (ed.) *Mind and language*. Oxford: Clarendon Press. 97–138.
- Dummett M (1976). 'What is a theory of meaning? (II)' In Evans G & McDowell J (eds.) *Truth and meaning: essays in semantics*. Oxford: Oxford University Press. 67–137.
- Glock H-J (2003). *Quine and Davidson on language, thought and reality*. Cambridge: Cambridge University Press. Chapters 5–8.
- Grandy R (1973). 'Reference, meaning and belief.' *The Journal of Philosophy* 70, 439–452.
- Heil J (1998). *Philosophy of mind: a contemporary introduction*. London: Routledge. Chapter 5.
- Hookway C (1988). *Quine: language, experience and reality*. Stanford: Stanford University Press. Chapters 8–10.
- MacIntyre A (1985). 'Relativism, power and philosophy.' *Proceedings and Addresses of the American Philosophical Association* 59, 5–22.
- Martin R M (1987). *The meaning of language*. Cambridge, MA: MIT Press. Chapter 22.
- Miller A (1998). *Philosophy of language*. Montreal: McGill-Queen's University Press. Chapters 4 and 8.
- Putnam H (1987). *The many faces of realism*. LaSalle, IL: Open Court. Lecture I.
- Quine W V (1960). *Word and object*. Cambridge, MA: MIT Press. Chapters I–II.
- Quine W V (1969). *Ontological relativity and other essays*. New York: Columbia University Press.
- Ramberg B T (1989). *Donald Davidson's philosophy of language: an introduction*. Oxford: Blackwell.
- Rudder-Baker L (1989). 'Instrumental intentionality.' *Philosophy of Science* 56, 303–316.
- Schiffer S (1989). *Remnants of meaning*. Cambridge, MA: MIT Press. Chapter 5.
- Tarski A (1949). 'The semantic conception of truth.' In Feigl H & Sellars W (eds.) *Readings in philosophical analysis*. New York: Appleton-Century-Crofts. 52–84.

Realism and Antirealism

K Green, Monash University, Melbourne, Australia

© 2006 Elsevier Ltd. All rights reserved.

Realism and antirealism in semantics can take a variety of forms, which are catalogued at the conclusion of this article. First, since realism looks like a metaphysical doctrine, it is necessary to set out how it relates to semantics. Some have claimed that since realism and antirealism concern metaphysics, they have little to do with semantics, truth, meaning, or

reference. Michael Devitt and Kim Sterelny argue that realism has two components: it is a doctrine about what exists and a doctrine about the nature of that existence. The realist about commonsense physical objects claims that they exist and that the nature of their existence is to exist independently of the mind. Devitt and Sterelny contrasted this realism with idealism: the view that while ordinary physical objects exist, their nature is to exist 'in the mind' in such a way that their existence is mind dependent (1999: 234–235). These doctrines, Devitt and

Sterelny argued, have nothing to do with semantics, for they can be stated without reference to semantic vocabulary such as 'meaning' or 'truth' (1999: 239). They position themselves as opposed to Michael Dummett, who has argued that the most illuminating way is to see them as implicitly hinging on differences in the theory of meaning (Dummett, 1978, 1991, 1992, 1993).

Realism does have two components, an existence component and one concerning the nature of that existence. Given this, antirealism, with regard to some range of entities, also has two forms. One denies existence. This is the antirealism of the atheist. The other denies mind independence, resulting in idealism and constructivism. In both cases, Dummett's contention, that the debate about what exists can be illuminated by seeing it as concerned with meaning and truth, appears correct. The traditional debate between idealists and realists arose from the skeptical questions posed by René Descartes in his *Meditations*. There Descartes wondered whether we know anything. Since in order to have knowledge one must have true beliefs, we could say that he wondered whether any of his beliefs were true. Descartes assumed that if his beliefs were true, they were made true by a reality that existed independently of him. Yet since he had access to that reality only through his 'ideas,' the question arose as to whether those ideas corresponded to anything, or were, like dreams, mere illusions. The idealists solved Descartes' radical skeptical problem by rejecting the gap between ideas and reality. The existence of the ideas themselves was taken to make our beliefs true. Formulated thus, the traditional distinction between realism and idealism can be seen to hinge on a dispute over what makes our beliefs true. Since it is all one whether we pose the skeptical problem directly for beliefs or think of it in terms of the sentences that express those beliefs, the metaphysical problem lends itself to a semantic formulation (Dummett, 1992: 132–134).

In his early writing, Dummett thought that the way to tackle the issue that divided realists from antirealists was to tackle the notion of truth directly, for he accepted, with Ludwig Wittgenstein, that the sentence is the smallest unit with which one can make a move in the language game and that therefore the notion of truth for sentences was primary (Dummett, 1973: 192–196). Looked at from the point of view of truth for sentences, the realist/antirealist debate hinges on whether truth is independent of us or is simply ideal warranted assertibility. A realist will say that truth is objective and independent of anything we do, but the price to pay for this is that there is no compelling reply to the skeptic. Also, if truth has nothing to do with our holding sentences

true, it can play no role in an account of the meaning of those sentences and is in danger of becoming an ungraspable we know not what. If, however, we want an account of truth that is central in a theory of meaning – which is a theory of what we know when we understand sentences – truth should be something we can grasp. Hence, Dummett analyzed truth in terms of warranted assertibility. As a result the principle of bivalence, which says that every meaningful assertoric sentence is either true or false, is given up. For there are sentences that are not justifiably assertible, yet it is not the case that we can confidently say that they are false, that is, that they will never be assertible. When the argument for antirealism directly concerns the notion of truth for sentences, it leads to questioning bivalence and the adoption of intuitionistic logic. But there are other arguments for antirealism that have a different character.

One variety of antirealism points to various speech acts that we can perform other than making truth-assessable statements. It claims that some sentences with which speakers appear to be making statements are not really descriptive utterances. So the possibility of a realist account of the truth of sentences for some area of discourse is excluded, because the discourse is not fact stating. The position of the expressivist in ethics, adopted by A. J. Ayer, is the paradigm of this form of antirealism (Ayer, 1946: 107). Like error theorists, who are discussed later in this article, expressivists suggest that we are in error about the actual semantics of our utterances. For them the realist's error lies in thinking that sentences of the disputed class describe facts.

Yet another issue divides realists about universals from nominalists. Like the realist about the physical world, the realist about universals at first appears to be making a metaphysical claim. It is that universals, as well as particulars, exist independently of us. As with realism about material objects, this claim can be seen to hinge on issues concerning that in virtue of which sentences are true. In order for the sentence 'The rose is red' to be true, one needs more than the mere existence of the particular rose being referred to. One needs it to be the case that the rose is red. If one understands this in terms of the rose instantiating the property of redness, one may be led to postulate the existence of universals, which are what general terms such as 'red' pick out. The nominalist, however, denies that one needs to explain the function of both singular and general terms by assigning a reference to them. According to Quine, for instance, singular terms refer, but predicates express their meanings; it is merely grammatical confusion that leads us to represent predication as a relation between two entities (Quine, 1970: 66–67).

The dispute over the existence of universals shows that not all forms of the realism/antirealism debate directly concern the notion of truth. What is at issue is often a question of the reference, or otherwise, of the subsentential parts of sentences. But these analyses flow naturally from questioning the nature of truth. Aristotle thought that both singular terms and general terms function by referring to, or picking out entities. Singular terms refer to objects; general terms refer to universals. On this model, a simple subject-predicate sentence is a string of names that is true just in case the object referred to by the singular term instantiates the universal referred to by the general term (Aristotle, 1984: 25–27). Bertrand Russell, puzzled by the meaningfulness of singular terms that fail of reference, proposed that in the vast majority of cases we can analyze names and definite descriptions, which appear to refer to objects, as really only involving descriptive phrases and quantifiers. ‘The rose on my table is red’ is then true just in case there is a unique thing that has the property of being a rose on my table as well as the property of being red (Russell, 1956). Dummett suggested that this is a move in the direction of antirealism (Dummett, 1991: 325). It is a semantic analysis that allows for the reduction of objects to bundles of properties.

Russell’s analysis of singular terms can be used to illustrate the way in which antirealism often takes a reductionist form. In 1904 Alexius Meinong proposed that ‘The golden mountain is golden’ is true in virtue of the subsistence of a nonexistent golden mountain to which the singular term ‘the golden mountain’ refers. Russell, objecting to a category of nonexistent existents, proposed that we reformulate the true sentence as saying that if anything is a golden mountain, it is golden. Russell was by no means an idealist, but in denying the existence of a referent of ‘the golden mountain,’ he went against the first component of a realist attitude to golden mountains. Similarly, the physicalist, being an antirealist about thoughts, will have to explain away the apparent truth of sentences such as ‘The terrifying thought stopped him in his tracks.’ A radical antirealist of the eliminativist slant might deem this sentence literally false. Alternatively, it could be treated as neither true nor false, because it is ill formed. But a reductionist antirealist with regard to thoughts will offer a paraphrase. This might take the form ‘Neurons playing the functional role of a warning of danger fired and stopped him in his tracks.’ Here the truth of the sentence involving an apparently nonreferring term is accounted for by offering a translation from which it is absent.

Both deeming some sentences neither true nor false because of failures of reference, and offering

reductionist paraphrases, constitute forms of antirealism of the broadly error-theoretic type. Error theory received its classic formulation in the work of Mackie (1977). The error theorist usually accepts that there is some range of sentences, typically sentences about physical things, for which the realist semantics holds. But the error theorist argues that we are led into error when we assume that the sentences of some other class share the same semantic structure. Mackie argued this for ethical statements. We are inclined to think of the semantic structure of the sentence ‘Charity is good’ as analogous to ‘Christ is good.’ If we accept the Aristotelian semantics sketched earlier in this article, this will involve thinking of charity as a kind of abstract object that instantiates the universal goodness. But these are queer kinds of entity. It is unclear how we have epistemological access to them, and it is arguable that we have been led to postulate their existence by an erroneous extension of the semantics of statements about physical objects. Adopting a version of this approach, Hartry Field argued, against the mathematical Platonist, that statements of mathematics that appear to involve truth grounded in reference to numbers really do no such thing and should not be deemed either true or false (Field, 1980: viii). When a person is persuaded on error-theoretic grounds that a realist semantics for some class of sentences is mistaken, but nevertheless wants to save the appearance of truth and falsehood for sentences of this class, he or she is characteristically forced to offer a reductionist account.

We can now construct a brief catalogue of forms of realism and antirealism. Global realism would treat all sentences realistically, and an ultra-global realist would accept not just classical logic but also the Aristotelian semantics according to which all singular terms refer to objects, while general terms refer to universals. Almost no philosopher is a global realist, and global antirealism is equally rare. Disputes arise in many areas. Platonists in mathematics oppose error theorists and constructivists; realists about mental entities oppose eliminativists and various forms of reductionist; realists about theoretical entities oppose instrumentalists; realists about commonsense material objects oppose phenomenologists and idealists; realists about possible worlds oppose ersatz theorists; realists in ethics oppose subjectivists, expressivists, and error theorists; realists about the past oppose presentists; and realists about the future oppose antirealists, for whom the future does not exist. Reductionist antirealism cannot be global, since it always involves a class of sentences interpreted realistically that provides a reduction class. Yet not all antirealism is reductionist. In the case of antirealism about the

future, there is a direct intuition that future-tense sentences are not now either true or false. Antirealists may deem some class of sentences to be nondescriptive, or descriptive but not involving a realist notion of truth. Among those who accept realist truth, there are those who analyze truth in terms of the existence of objects and universals, and others who take an antirealist attitude toward either objects or universals.

See also: Descriptions, Definite and Indefinite: Philosophical Aspects; Empty Names; Nominalism; Objectivity in Moral Discourse; Reference: Philosophical Theories; Semantic Value; Verificationism.

Bibliography

- Aristotle (1984). 'De interpretatione.' In Barnes J (ed.) *The works of Aristotle*. Princeton: Princeton University Press. 1, 25–27.
- Ayer A J (1946). *Language, truth and logic*. London: Victor Gollancz.
- Blackburn S (1993). *Essays in quasi-realism*. New York: Oxford University Press.
- Devitt M & Sterelny K (1999). *Language and reality: an introduction to the philosophy of language*. Oxford: Basil Blackwell.
- Dummett M (1973). *Frege: philosophy of language*. London: Duckworth.
- Dummett M (1978). 'Realism.' In *Truth and other enigmas*. London: Duckworth. 145–165.
- Dummett M (1991). *The logical basis of metaphysics*. Cambridge, MA: Harvard University Press.
- Dummett M (1992). 'The metaphysics of verificationism.' In Hahn L E (ed.) *The philosophy of A. J. Ayer*. La Salle, IL: Open Court. 129–148.
- Dummett M (1993). 'Realism and anti-realism.' In *The seas of language*. Oxford: Oxford University Press. 462–478.
- Field H (1980). *Science without numbers*. Princeton: Princeton University Press.
- Mackie J L (1977). *Ethics: inventing right and wrong*. Harmondsworth: Penguin.
- Quine W V O (1970). *Philosophy of logic*. Englewood Cliffs: Prentice Hall.
- Russell B (1956). 'On denoting.' In Marsh R (ed.) *Logic and knowledge*. London: George Allen and Unwin. 41–56.
- Wright C (1993). *Realism, meaning and truth*. Oxford: Blackwell.

Reference: Philosophical Theories

A Sullivan, Memorial University of Newfoundland, St. John's NL, Canada

© 2006 Elsevier Ltd. All rights reserved.

What Is Reference?

Reference is the relation that obtains between a use of a linguistic expression and what it stands for or denotes. Philosophical theories of reference aim to explain this relation. They address questions such as: Can we specify necessary or sufficient conditions for a use of an expression to single out some particular referent? What makes it the case that a sequence of marks or noises refers to what it does? What grounds the conventional word-referent relations that we take for granted in the unreflective course of linguistic communication?

Reference is a fundamental concept within the sub-discipline of philosophical inquiry that studies language and thought. It is intimately tied to some other core concepts, such as meaning, truth, and representation, and it lies at the root of various general philosophical issues. To cite a few illustrative examples (some of which are discussed later): [1] there is a venerable tradition of using data about referring expressions to support metaphysical and epistemological conclusions about their referents, running at

least from Aristotle's *Categories* through to many influential 20th-century works; [2] the ancient problem of nonbeing – cf. Plato's *Parmenides* for a classic discussion, and Lewis (1986) for discussion of some contemporary variants – is in part a problem about reference to nonactual, or nonconcrete, objects; [3] reference plays a central role in some of the most provocative and broad-ranging recent work in philosophy, including in particular the varieties of externalist arguments spurred by Kripke (1972) and Putnam (1975).

Proper names take center stage in debates about reference, as names are explicitly introduced to refer to specific individuals. Other much-discussed, basic cases of reference include uses of demonstratives (such as 'that duck') and indexical pronouns ('she') to refer to individuals which the speaker currently has in mind. Beyond these paradigm cases, though, there is not much uncontested ground. For example, there are long-standing debates over whether definite descriptions – expressions of the form 'the F' – should be classified as referring expressions (for discussion, *see Descriptions, Definite and Indefinite: Philosophical Aspects*).

In addition to such differences of opinion, there also are significant differences of usage. In its strictest sense, 'reference' applies only to the relation between singular terms (such as names and pronouns) and

what they are used to single out. On this usage, 'reference' names the simplest sort of word-object relation; and whatever the semantic relations are, precisely, that are involved in, say, general terms ('justice') or predicates ('is tall'), they are distinct from reference. In a broader sense of the term 'reference,' any expression that makes a difference to the truth-conditions of sentences in which it occurs thereby has a referent – general terms refer to universals, predicates refer to properties, and so on. So, in the former, strict sense of the term, the theory of reference is more or less co-extensive with the study of the semantics of singular terms, whereas, in the latter, broad sense, the business of the theory of reference is to assign semantic values to all independently meaningful expressions.

There also are differing views as to whether the term 'reference' just applies to language-world relations, or whether mental phenomena, such as concepts, also should be said to refer. On the one hand, some treat 'reference' as a two-place relation between a representation and what it represents, and so take the term to apply in kind to (at least) word-world and concept-world relations. In this inclusive sense, all thought or talk about something in particular involves reference to it. On the other hand, there are those who hold that there are important differences between referring to something and thinking of it, and so restrict to the term 'reference' to language-world relations. Their point is not so much to deny that the problems of word-reference and of concept-reference are intimately related – that is true on virtually any approach to the relation between language and thought. Rather, their point is that referring is something that speakers do in the course of performing speech acts, not something that words themselves do. These theorists hold that the fundamental notion in the theory of reference is *speaker's reference* – an irreducibly four-place relation involving a speaker, an expression, an audience, and a context.

For comprehensiveness, I will use the term 'reference' in a fairly broad, inclusive sense. For brevity, I will concentrate on uses of words, and I will treat reference as a two-place relation between a use of a word and its referent. Given these preliminaries and assumptions, the central problem for a theory of reference is to come up with a true, informative account of what the 'R' stands for in the schema:

[Ref] 'X' refers to Y if 'X' stands in relation R to Y

Descriptivism

Descriptivist Theories of Reference

Even though philosophical problems that turn on the notion of reference are more or less as old as

philosophy, the development of theories of reference – of explicit attempts to articulate how it is that terms reach out to their referents – does not occur until the seminal semantic work of Frege and Russell. (Some might argue that the idea theorists of the early Modern period had a theory of reference – see **Ideational Theories of Meaning** for discussion.) Down different avenues, Frege and Russell both defend descriptivist views about reference, and their views dominated the scene for most of the 20th century.

Descriptivists take the definite description ('the tallest woman in Mongolia', 'the millionth car to roll off this assembly line') as the paradigm for understanding the word-object referential link. The relation between descriptions and their referents is relatively straightforward: the term expresses an identifying condition, and designates whatever (if anything) uniquely satisfies that condition. In its most general sense, 'descriptivism' names the view that this is how reference works, that words refer to specific things by expressing an identifying condition which singles out the referent. For starters, then, let us characterize the core tenet of a descriptivist theory of reference as follows:

[D] 'X' refers to Y if Y satisfies the descriptive condition expressed by 'X'

There are more specific variants of descriptivism, which just apply to a distinctive category of terms, as opposed to making claims about reference in general. (For example, Lewis (1970) develops a descriptivist approach to theoretical terms.) The most common usage of 'descriptivism,' though, specifically designates an approach to ordinary proper names. According to this view, names such as 'Aristotle' or 'Romania' hook onto and express information about something in particular, akin to descriptions, by expressing a uniquely identifying condition and designating whatever satisfies the condition. This approach to ordinary proper names is the kernel in common among Frege's and Russell's otherwise rather divergent views about reference.

Frege's and Russell's Versions of Descriptivism

In their earliest work, Frege and Russell both espoused referentialist views of meaning, according to which the meaning of an expression is whatever it is to which the expression refers. However, they came to believe that referentialism must be rejected or modified, because two terms can have the same referent but differ in meaning.

Frege's signature example concerns the names 'Hesperus' and 'Phosphorus.' According to the lore, the Evening Star (the brightest heavenly body in the evening sky at certain times of the year) was baptized

‘Hesperus,’ whereas the Morning Star (the brightest body in the dawn sky at certain times) was baptized ‘Phosphorus’. With the growth of astronomical knowledge, it turns out that Hesperus *is* Phosphorus, and that is not a star at all but the planet Venus. Frege and Russell agree that pairs of sentences such as the following differ in meaning:

- 1a. Hesperus is Hesperus.
- 1b. Hesperus is Phosphorus.
- 2a. The ancient Babylonians believed that Hesperus is visible in the evening.
- 2b. The ancient Babylonians believed that Phosphorus is visible in the evening.

However, because these sentences are constructed from co-referential parts, the referentialist view is committed to treating them as synonymous.

Frege (1892) concludes that, in general, the meaning of a term is not just its referent. Instead, every significant linguistic expression is associated with a sense that determines its referent. The sense specifies the condition that must be met in order to be the referent, which clearly differ for the cases of ‘Hesperus’ and ‘Phosphorus.’ Co-referential expressions can have distinct senses, and expressions need not have an actual concrete referent to have a sense. Thus, on Frege’s view:

[D, Frege] ‘X’ refers to Y if Y satisfies the sense ‘X’ expresses

This type of view is called ‘mediated’ or ‘indirect’ reference. The link between word and referent is *indirect* because there is something *mediating* the word-referent link. On this approach, any referential relation involves three terms: [1] a sign or symbol, [2] its sense, and [3] the referent picked out by the sense. (Note that there may be no actual entity to serve as term [3] – as in the case of, say, ‘Santa Claus’ or ‘Pegasus.’ However, such cases still fit with the general picture, as the names express a sense that specifies the condition that must be met to count as the referent.) (see **Sense and Reference: Philosophical Aspects** for more on Frege’s view.)

Russell rejects Frege’s views, arguing that the postulation of intermediaries between word and referent creates rather than solves problems. Russell’s view is premised on a firm distinction between the following two sorts of term: “(1) a *name*, which is a simple symbol, directly designating an individual which is its meaning, and having this meaning ... independently of all other words; (2) a *description*, which consists of several words, whose meanings are already fixed, and from which results whatever is to be taken as the ‘meaning’ of the description” (1919: 283). (Following Russell [1918], I’ll call the former

‘logically proper names.’) Concerning logically proper names, Russell is a direct reference theorist, an unrepentant referentialist. Reference is *direct* in that there is no third term mediating the word-referent relation. The meaning of a logically proper name is its referent.

However, for the reasons illustrated by (1a–b) and (2a–b) above, Russell argues that ordinary proper names (such as ‘Hesperus’ or ‘Aristotle’) belong in the category of descriptions, not of logically proper names. He holds that it is not possible for co-referential logically proper names to differ in meaning – “For the name itself is merely a means of pointing to the thing ... so that if one thing has two names, you make exactly the same assertion whichever of the two names you use ...” (1918: 245). In contrast, descriptions (such as ‘the inventor of bifocals’ and ‘the first Postmaster General of the US’) can differ in meaning while describing the same referent. Russell concludes that, since ordinary proper names are subject to the phenomena illustrated by (1a–b) and (2a–b), they are really just disguised or abbreviated descriptions – “Common words, even proper names, are usually really descriptions” (1911: 253). Thus, Russell also espouses a form of descriptivism:

[D, Russell] For any ordinary proper name ‘N’, ‘N’ refers to Y if Y satisfies the description ‘the F’ which the speaker has in mind in uttering ‘N’

Differences Between Descriptivist Views

Despite the fact that Frege and Russell both espouse descriptivist accounts of ordinary names, their approaches to reference are quite far apart. On Russell’s direct reference view, there must be some primitive cases in which it is impossible to have sameness of referent but distinctness of meaning. He holds that we must be directly acquainted, and not just indirectly linked via descriptions, with the referents of some of our terms. (Otherwise “... we get the conclusion that we know nothing at all about ...” the referents of our terms, as Russell puts the point in a letter to Frege [reprinted in Frege (1980: 169)].) So, apart from, and more fundamental than, his views about ordinary names, Russell espouses an acquaintance-based theory of reference:

[A] ‘X’ refers to Y if Y is the entity with which the speaker is acquainted that the speaker intends to single out by uttering ‘X’

(It is crucial to Russell’s view that acquaintance is *immediate*, in that for any co-referential pair ‘X’-‘X*’, one cannot be acquainted with their referent and yet judge that X is distinct from X*.) Russell holds that one can only use a logically proper name

to refer to something to which one is directly acquainted (1918: 201). Furthermore, even in the case of a description, one can only understand a description if one is acquainted with the meanings of the terms of which it is composed (1911: *passim*) – i.e., to understand the term ‘the tallest woman in the room’, one must be acquainted with the meanings of ‘tallest,’ ‘woman,’ ‘room,’ and so on.

According to Russell, one needs to fall back on something like [A] in order to address questions that are subsequent to, and more fundamental than, a commitment to some version of [D]. In effect, [D, Russell] reduces the ordinary-naming-relation to the description-relation; but still, according to Russell, we do not yet have a theory of reference. In order to explain how ‘the tallest woman in the room’ expresses the specific identifying condition that it does, one needs an account of the referents of its primitive parts – i.e., a theory of reference for ‘tallest,’ ‘woman,’ and so on. Furthermore, this account had better not just appeal to more descriptions, or else we are not getting anywhere. So, Russell holds that [D] needs to be supplemented, if we are to have an account of how our utterances get hooked up to particular referents.

This problem never seems to have bothered Frege; and yet it is not clear whether this is an oversight. (Perhaps Frege’s senses are relevantly different from Russell’s descriptions; perhaps the moral is that a direct-reference-descriptivism needs to be supplemented by something like [A], whereas an indirect-reference-descriptivism does not.) In any case, it can be argued that [D] is not yet a theory of reference, because it presupposes, rather than explains or grounds, reference. To help oneself to some specific Y as that which satisfies the descriptive condition expressed by ‘X,’ so this Russellian allegation goes, is to beg the very question that a theory of reference ought to answer.

No such worry applies to [A]. Acquaintance is a clear and distinct answer to the fundamental question of reference. However, it puts rather tight constraints on reference. By 1918, Russell’s insistence that one must be immediately acquainted with something in order to refer to it lead him to the view that only ‘this,’ used to refer to current sense impressions, qualifies as a referring expression. Subsequent theorists have found this unacceptable, and so reject Russell’s tight restrictions on reference.

For decades, work in the theory of reference consisted largely of the development of other variants of descriptivism (cf. Searle [1958]; Strawson [1959]). However, some deep criticisms of descriptivism emerged in the 1970s, which lead to another major wave of original work in the theory of reference.

Antidescriptivism and the Causal-Historical Theory of Reference

Problems With Descriptivism

Kripke (1972) is the most thorough and influential critic of descriptivism, whereas Donnellan (1970), Putnam (1975), and Kaplan (1977) also develop important criticisms. Three widely discussed problems with descriptivist theories of names are known as the modal problem, the epistemic problem, and the semantic problem.

To illustrate the modal problem, take any proper name (say, ‘Aristotle’) along with descriptions that are true of its referent (‘the most famous of Plato’s students,’ or ‘the teacher of Alexander’). Consider the differences in truth-conditions of the following:

1. Aristotle had brown hair.
2. The most famous of Plato’s students had brown hair.

Even if (1)–(2) actually agree in truth-value, they need not have. Had things gone differently (and, say, Aristotle died young, or went into commerce), then the truth-values of (1)–(2) could vary independently – i.e., each might have been true while the other is false. A similar contrast could be drawn using virtually any name-description pair, and many take this phenomenon to indicate a deep difference between names and descriptions.

To illustrate the epistemic problem, if ‘Aristotle’ meant something like ‘Plato’s most famous student,’ then the following would be self-evident to anyone who understands it:

3. Aristotle is Plato’s most famous student.

This consequence is false, as (3) is not at all trivial or self-evident. Again the point generalizes, having nothing to do with this specific example; and the putative upshot is that names like ‘Aristotle’ are not equivalent in meaning to any particular descriptions.

The semantic problem is that speakers may be competent to use a name even if the descriptive conditions they associate with the name does not suffice to single out the referent. That is, one might just think that ‘Aristotle’ names some famous Greek guy (or perhaps even confuse Aristotle with Heraclitus), but yet nonetheless one could still refer to Aristotle by uttering ‘Aristotle’. Again, the conclusion seems to be that there are important differences between reference by name and the satisfaction of descriptive conditions. (*see Proper Names: Philosophical Aspects for discussion.*)

Collectively, these three problems are widely thought to spell the end for descriptivist accounts of names. (This is not unanimous, however – cf. Searle

[1983] and Bach [1987] for descriptivist replies.) Kripke (1972) and Putnam (1975) develop similar problems for some varieties of general terms. Still other, related problems for descriptivism are raised by indexicals – even if satisfactory descriptivist accounts could be given for some terms, it is not plausible to hold that all reference is descriptive, as is shown by cases like ‘It is quiet here now’ (cf. Kaplan [1977]).

The Causal-Historical Theory of Reference

The critics of descriptivism put forth a different picture of reference, known as the causal-historical theory. Originally, the theory is only developed for proper names (cf. Donnellan [1970]; Kripke [1972]) and natural kind terms (cf. Kripke [1972]; Putnam [1975]). However, there have been attempts to fashion more general, comprehensive views of reference along these lines (for discussion, see Stalnaker [1997]; Devitt and Sterenly [1999]).

The crux of this picture is that certain expressions refer to certain things in virtue of a causal-historical relation between word and object, initially fixed during a dubbing or baptism and propagated from there to subsequent speakers, who implicitly defer to that initial dubbing in using the expression to refer. The view has it that descriptivists are wrong to demand that, in order to significantly use a term, speakers are required to have an identifying description of its referent. Rather, once a word-referent convention is in place, all that is required in order to use the term to refer is a deferential intention to comply with the convention – i.e., the (implicit) intention to use ‘X’ to refer to what others have used ‘X’ to refer to. On this approach, reference depends not simply on what the speaker thinks, but also on certain factors external to the speaker’s head, factors pertaining to the speaker’s linguistic community and to the environment in which the expression in question evolved. The basic idea is:

[CHT] ‘X’ refers to Y if ‘X’ stands in the right sort of causal-historical relation to Y

(see **Causal Theories of Reference and Meaning** for more detail).

Many of Russell’s successors have worked to loosen the tight restrictions on his notion of reference, by replacing acquaintance with something more broadly applicable and accessible. The [CHT] is a significant step forward in this direction. It is an externalist theory of reference. Internalist views of reference hold that reference is fully determined by the intrinsic content of the speaker’s head; externalists deny this. Externalists hold that two speakers could be in the same mental state and utter the same

sounds, but nonetheless their utterances have different referents – because of differences in the histories of how the words reached them, or of differences in the environments to which they are causally connected. (For discussion, see **Externalism about Content**.)

Intuitively, reference is a relation involving (at least) a speaker, an expression, and a referent; externalists hold that it is a mistake to think that the nature of this relation could be fully characterized by analyzing only one term of this relation (i.e., the intrinsic content of the speaker’s head). Most theorists who articulated views about reference prior to the [CHT] (including Frege and Russell) are commonly characterized as internalists. In the wake of the [CHT], though, externalism about reference has become the norm. Most subsequent theorists agree that the [CHT] makes a progressive step forward, in giving some of the weight in determining reference to the extrinsic relations in which speakers stand to other speakers and to their environments.

The [CHT] led to a resurgence of interest in the notion of direct reference. Many take the arguments against descriptivism to apply to all indirect reference theories, and so to make it plausible that the meaning of a term is just its referent. In addition, many see a complementary fit between the causal-historical chain of transmission story about reference and the direct reference approach to the question of meaning or content. (see **Proper Names: Philosophical Aspects** for discussion.)

Problems With the Causal-Historical Theory

There are many problems with the causal-historical theory of reference. For instance, Evans (1973) points out that some words have in fact changed their referents over time, even despite the deferential intention to preserve the convention in question, and argues that the [CHT] cannot account for this phenomenon. For another example, Searle (1983: 249) offers as a *reductio* of the theory the putative consequence that his utterances of ‘Aristotle’ might refer to a barstool in Hoboken “if that is what the causal chain happened to lead to.” To cope with these and other problems, many have sought to defend hybrid views, which aim to combine the strengths while avoiding the weaknesses of [D] and [CHT] (cf. Evans [1973, 1982]).

More generally, though, the [CHT] leaves important questions unaddressed. One problem concerns specifying exactly what should count as ‘the right sort of causal-historical relation’. As Stalnaker (1997: 543) puts it: “Causal connections are ubiquitous ... there are a great many individuals causally

implicated in [any] use of [any] name. . . . A proper causal theory of reference would have to specify just what sort of causal connection is necessary and sufficient for reference, and that is a notoriously difficult demand.” Kripke (1972: 96) is explicit that he is not offering a reductive analysis of reference: “When the name is ‘passed from link to link’, the receiver of the name must, I think, intend to use it with the same reference as the man from whom he heard it, . . . [T]he preceding account hardly *eliminates* the notion of reference; on the contrary, it takes the notion of intending to use the same reference as a given.” (Cf. Kaplan’s [1990] discussion of the point that the intention to preserve reference is not itself a causal notion.) Here we see why the [CHT] is open to the allegation that it begs the key question in the theory of reference, by helping itself to successful reference. Critics allege that it boils down to the relatively toothless claim that: ‘X’ refers to Y if everyone else has used ‘X’ to refer to Y.

Of course, the [CHT]’s developers recognize the need for a separate story about what determines reference in the first link in a causal-historical chain of transmission. However, their remarks about baptisms or dubbings fall far short of a comprehensive account. (Cf. Devitt and Sterenly’s [1999] discussions of the ‘qua’-problem for a sense of some of the relevant unfinished business.) So, while most agree that the [CHT] provides a better picture than [D], it has yet to be fleshed out into a specific, comprehensive theory of reference.

Skepticism, Naturalism, and Minimalism About Reference

Some influential thinkers are skeptical about the prospects of coming up with a satisfactory theory of reference. Chomsky (1993), for instance, argues that the debates about reference are hopelessly flawed because they presuppose an implausible, unscientific conception of language. The precise import of Chomsky’s complaints is unclear, though. In particular, and explicitly, they do not apply to the relatively concrete and tractable notion of speaker’s reference (i.e., a four-place relation between speaker, expression, audience, and context).

Quine (1960) is perhaps the most influential skeptic about reference, arguing that there are no determinate facts about reference, no objective thing to be gotten right. Quine takes this conclusion (which he calls the ‘inscrutability of reference’) to be an instance of the general point that theories are underdetermined by evidence – i.e., for any finite set of data, there are an infinite number of distinct theories that are consistent with, and could purport to explain, the data.

In the case of reference, the claim is that, for any finite set of evidence about a given speaker, there are a number of distinct word-referent assignments that are compatible with it. (To take one of Quine’s central examples, no data will tell us whether a given speaker is using ‘rabbit’ to refer to rabbits, to time-slices of rabbits, to undetached rabbit parts, and so on.) There is no principled way of deciding between these competing assignments; and hence, there is no determinate fact of the matter as to what words are used to refer to. From here, Quine draws pessimistic conclusions about the scientific credentials of such notions as reference.

Many aspects of Quine’s approach have been criticized. Still, much subsequent work on reference is addressed to Quine’s underlying challenge of explaining reference within the ontology and methodology of the natural sciences. This is known as the demand for a naturalistic theory of reference, and a number of candidates were proposed in recent decades.

One such candidate is another causal theory of reference – also called the ‘information-theoretic’ approach (cf. Stampe [1977]; Dretske [1981]; Fodor [1990]). The general idea is that information is a species causal co-variance, that for a word ‘X’ to mean something Y is for tokens of ‘X’ to reliably indicate Y. As applied to reference, the view comes to something like:

[CT] ‘X’ refers to Y if Y tends to cause or bring about tokens of ‘X’

However, in addition to inheriting some of the general worries about [CHT], [CT] is subject to counterexamples. For almost any word-referent pair ‘X’–Y, it is not difficult to find conditions in which things distinct from Y tend to cause ‘X’s, and conditions in which Y does not tend to cause ‘X’s. Despite ingenious work to patch things up, the consensus seems to be that this type of causal theory can only succeed in delivering an account of reference that accommodates our intuitions about the normativity and determinacy of reference by smuggling in intentional, semantic notions, and thus by helping itself to reference and forgoing naturalism. (For discussion see Loewer (1997), and **Causal Theories of Reference and Meaning.**)

Another naturalistic approach to reference is the teleological theory, whose most thorough proponent is Millikan (1984). The idea here is that:

[T] ‘X’ refers to Y if it is the function of ‘X’s to indicate Y

Here ‘function’ is a familiar notion from evolutionary biology – roughly, X’s function is what X does, that explains why X exists. However, this theory, too,

has serious troubles accounting for the normativity and determinacy of reference (cf. Fodor [1990]; Loewer [1997]). Although teleological considerations afford a compelling story about why agents like us have developed the capacity to refer, it is a long way from there to specifying necessary or sufficient conditions for determining the referent of a given use of a word.

Another reaction to this long and sordid history is the minimalist (or ‘deflationary’) approach (cf. Horwich [1998]; Field [2001]). Minimalists hold that reference is simple, not reducible to anything more fundamental. (Minimalism about reference is often conjoined with minimalism about truth – see **Truth: Theories of in Philosophy** for discussion.) Our meaning what we do by ‘refers’ consists entirely in our inclination to accept disquotational schemas – ‘Aristotle’ refers to Aristotle, ‘Romania’ refers to Romania, and so on – and there is nothing else to be said, by way of necessary or sufficient conditions. Rather, all that should be said is something like:

[M] ‘X’ refers to X

[M] is of course no definition: reference is not reduced to, or explained in terms of, anything more basic. However, one conclusion that is fairly well supported by the failed attempts to define reference is that such a reductive analysis may well be unattainable. To some extent, it is a matter of taste whether one classifies minimalist views as theories of reference or as a variety of skepticism about reference.

Even if there is something right about the minimalist approach, though, [M] hardly seems to illuminate the content of the concept ‘reference.’ Furthermore, there are specific objections to minimalism (cf. Loar [1995]; Loewer [forthcoming]). For example, it is not clear that there could be a satisfactory minimalist account of indexical expressions. (That is, the disquotational line on ‘Aristotle’ and ‘Romania’ does not carry over smoothly to ‘I’ or ‘here.’) Because these are core, basic cases of reference, this might point to something deeply suspect about minimalist views.

Summary

The fundamental problem in the theory of reference (given our preliminary assumptions) is to come up with a true, informative account of what the ‘R’ stands for in the schema:

[Ref] ‘X’ refers to Y if ‘X’ stands in relation R to Y

Work in the theory of reference largely consists of disputes between two camps, descriptivists and causal-historical theorists:

[D] ‘X’ refers to Y if Y satisfies the descriptive condition expressed by ‘X’

[CHT] ‘X’ refers to Y if ‘X’ stands in the right sort of causal-historical relation to Y

Descriptivists tend to be internalists, and descriptive reference is mediated or indirect, whereas the causal-historical theory is externalist and more friendly to direct reference. As we have seen, though, whereas [D] and [CHT] are major players in the semantics of proper names, both are – albeit from different directions – open to the allegation that they presuppose rather than explain or ground reference.

Bracketing off reference via descriptive conditions and reference via deference to causal-historical chains (as well as the hybrid views), there are at least four other basic answers to the fundamental question of reference (respectively, acquaintance, causation, teleology, and minimalism):

[A] ‘X’ refers to Y if Y is the entity with which the speaker is acquainted that the speaker intends to single out by uttering ‘X’

[CT] ‘X’ refers to Y if Y tends to cause or bring about tokens of ‘X’

[T] ‘X’ refers to Y if it is the function of ‘X’s to indicate Y

[M] ‘X’ refers to X

The problem with Russell’s version of [A] is that it is incompatible with the platitude that reference is a tool used in the intersubjective communication of information – i.e., acquaintance is a private subjective affair, but reference is not. However, there is a recent resurgence of interest in the notion of acquaintance, and it promises to avoid some of these problems. [CT] and [T] have certainly not been ruled out, but counterexamples have yet to be satisfactorily addressed. (Note that it is indicative that there are two causal theories on this short-list – most philosophers concede that causal relations between speakers and their environments have to play a fundamental role in a satisfactory theory of reference.) Proponents of [M] also have some specific problems to account for, over and above its general tendency to elicit the atavistic impulse that the question of reference has not really been addressed.

Much philosophical work continues on these problems, and these debates are not in danger of conclusive resolution any time soon.

See also: Causal Theories of Reference and Meaning; Descriptions, Definite and Indefinite: Philosophical Aspects; Direct Reference; Empty Names; Externalism about Content; Ideational Theories of Meaning; Proper Names: Philosophical Aspects; Saussure: Theory of the Sign; Sense and Reference: Philosophical Aspects; Truth: Theories of in Philosophy.

Bibliography

- Bach K (1987). *Thought and reference*. Oxford: Oxford University Press.
- Chomsky N (1993). *Language and thought*. Wakefield: Moyer Bell.
- Devitt M & Sterenly K (1999). *Language and reality* (2nd edn.). Cambridge, MA: MIT Press.
- Donnellan K (1970). 'Proper names and identifying descriptions.' *Synthese* 21, 256–280.
- Dretske F (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Evans G (1973). 'The causal theory of names.' *Proceedings of the Aristotelian Society* 47, 187–208.
- Evans G (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Field H (2001). *Truth and the absence of fact*. Oxford: Oxford University Press.
- Fodor J (1990). *'A theory of content' and other essays*. Cambridge, MA: MIT Press.
- Frege G (1892). 'On sense and reference.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 175–192.
- Frege G (1980). *Philosophical and mathematical correspondence*. Kaal H (trans.). Oxford: Blackwell.
- Horwich P (1998). *Meaning*. Oxford: Oxford University Press.
- Kaplan D (1977). 'Demonstratives.' In Almog J, Perry J & Wettstein H (eds.) (1989). *Themes from Kaplan*. Oxford: Oxford University Press. 481–564.
- Kaplan D (1990). 'Words.' *Proceedings of the Aristotelian Society* 64, 93–120.
- Kripke S (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Lewis D (1970). 'How to define theoretical terms.' *Journal of Philosophy* 67, 427–446.
- Lewis D (1986). *On the plurality of worlds*. Oxford: Blackwell.
- Loar B (1995). 'Reference from the first-person perspective.' In Villanueva E (ed.) *Contents*. Astacadero: Ridgeview. 55–76.
- Loewer B (1997). 'A guide to naturalizing semantics.' In Hale B & Wright C (eds.) *A companion to the philosophy of language*. Oxford: Blackwell. 108–126.
- Loewer B (forthcoming). Comments on Field's 'Truth in the absence of fact.' *Philosophical Studies*.
- Millikan R (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
- Putnam H (1975). 'The meaning of "meaning."' In Gunderson K (ed.) *Mind, Language, and Reality*. Cambridge: Cambridge University Press. 131–193.
- Quine W V (1960). *Word and object*. Cambridge, MA: MIT Press.
- Russell B (1911). 'Knowledge by acquaintance and knowledge by description.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 249–262.
- Russell B (1918). 'The philosophy of logical atomism.' In Marsh R C (ed.) *Logic and Knowledge*. London: Unwin Hyman, 1956.
- Russell B (1919). 'Descriptions.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 279–287.
- Searle J (1958). 'Proper names.' *Mind* 67, 166–173.
- Searle J (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Stalnaker R (1997). 'Reference and necessity.' In Hale B & Wright C (eds.) *A companion to the philosophy of language*. Oxford: Blackwell. 534–553.
- Stampe D (1977). 'Toward a causal theory of linguistic representation.' *Midwest Studies in Philosophy* 2, 42–63.
- Strawson P (1959). *Individuals*. London: Methuen.

Referential versus Attributive

A Barber, The Open University, Milton Keynes, UK

© 2006 Elsevier Ltd. All rights reserved.

Donnellan's Contrast

Consider two utterances of the same sentence, S:

S: Smith's murderer is insane.

The first utterance, U(S)_{ATT}, is prompted by the speaker's discovery of Smith's savagely mutilated corpse roasting on a spit, with no culprit in evidence. The second, U(S)_{REF}, is prompted by an observation years later of a defendant, Jones, seen cross-eyed and frothing at the mouth as he climbs into the dock on trial for Smith's murder. As it happens, the real killer, Robinson, goes free and is perfectly sane (his crime

notwithstanding) while Jones, though falsely charged, is as mad as he looks. Many find it natural to suppose that the speaker of U(S)_{ATT} has said nothing correct about anyone while the speaker of U(S)_{REF} has said something correct about Jones, even if perhaps she has used the wrong words to do so. This is an example of the contrast between what Keith Donnellan labels – misleadingly on some views – attributive and referential uses of definite descriptions. Here is Donnellan's own characterization of the contrast:

A speaker who uses a definite description attributively in an assertion states something about whoever or whatever is the so-and-so. A speaker who uses a definite description referentially in an assertion ... uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or

thing [T]he definite description is merely one tool for doing a certain job – calling attention to a person or thing – and in general any other device for doing the same job, another description or a name, would do as well. In the attributive use, the attribute of being the so-and-so is all important, which it is not in the referential use (Donnellan, 1966a: 285).

Attempts to account for how the contrast arises lie at the heart of work on the nature of the semantics/pragmatics boundary. This entry looks at the origins of Donnellan's distinction in his criticism of Russell's theory of descriptions, and then considers two approaches to the contrast, one that treats it as pragmatic in origin and the other that treats it as a genuinely semantic distinction.

Donnellan's Use of the Contrast against Russell

Donnellan introduced the distinction with a view to undermining Russell's theory of definite descriptions (Russell, 1905, 1919). Russell argued that the surface form of (1) below is misleading. It has an underlying logical form that, when interpreted, delivers the truth condition set out in (2):

- (1) 'The fish jumped'
- (2) An utterance of the sentence in (1) is true if and only if $\exists x ((x \text{ is a fish} \ \& \ \forall y (y \text{ is a fish} \rightarrow x = y)) \ \& \ x \text{ jumped})$

Definite descriptions, in other words, are non-namelike: nothing in the underlying logical form of sentences containing them corresponds to an object referred to by 'The fish.' (Russell also held that ordinary names are non-namelike in this respect too, but that is another matter.) By denying that the semantic function of definite descriptions is to pick out a single object in order then to say something about it, Russell hoped to solve a variety of puzzles. For example, it became possible to explain how a definite description can be used meaningfully even when less than one (as in (3)–(4)) or more than one (as in (5)) contextually salient object satisfies the description; or how identity statements made using definite descriptions can be informative as in (6) when an equivalent name would not be (as in (7)).

- (3) 'The king of France is bald': false but meaningful.
- (4) 'The king of France does not exist': meaningful and true.
- (5) 'The woman in the magician's box has been sawn in half': false but meaningful.
- (6) 'The author of the Waverley novels is Walter Scott': true and informative.
- (7) 'Walter Scott is Walter Scott': true but uninformative.

From Russell's perspective, these cases are puzzling only if one assumes that the entire semantic function of a phrase of the form 'the ...' is to single out an object that uniquely satisfies the property expressed by '...', in order then to say something about it with the rest of the sentence. If truth conditions are instead assigned along the pattern in (2), these examples cease to be troublesome.

According to Donnellan (1966a), Russell's theory copes adequately with $U(S)_{ATT}$ but not with $U(S)_{REF}$. It assigns the same truth conditions to both utterances:

An utterance of *S* is true if, and only if, $\exists x ((x \text{ murdered Smith} \ \& \ \forall y (y \text{ murdered Smith} \rightarrow x = y)) \ \& \ x \text{ is insane})$

Intuitively, $U(S)_{REF}$ would need to be withdrawn if Jones were in fact sane, while $U(S)_{ATT}$ would not be affected; and Robinson's mental health is irrelevant to $U(S)_{REF}$ despite being highly relevant to $U(S)_{ATT}$. Donnellan took these intuitions, and others elicited by similar cases (e.g. 'Who is the man drinking a martini?' asked of someone who is in fact drinking water) to refute Russell's theory, at least in so far as the latter was intended to give a comprehensive account of definite descriptions.

It is unclear whether Donnellan thought his distinction had a pragmatic or a semantic origin. The fact that he takes it to refute Russell suggests he took it to be semantic. Russell's theory was, after all, a theory of the semantics of definite descriptions, not of their pragmatics. But he persistently talks of *uses* of definite descriptions, and at one point even writes explicitly that

[...the] grammatical structure of the sentence seems to me to be the same whether the description is used referentially or attributively: that is, it is not syntactically ambiguous. Nor does it seem at all attractive to suppose an ambiguity in the meaning of the words; it does not appear to be semantically ambiguous. (Perhaps we could say that the sentence is pragmatically ambiguous: the distinction between roles that the description plays is a function of the speaker's intentions.) (Donnellan, 1966a: 297)

It is possible that Donnellan's position, further developed in Donnellan (1968), is just that no pragmatic theory is available to complement Russell's semantic theory in such a way as to explain referential uses as well as attributive ones. In that case, even if there is no semantic ambiguity, Russell's semantic theory would have to be rejected. But a more likely explanation of the contrary indicators in his paper is the primitive understanding of the semantics/pragmatics boundary at the time.

Whatever we take Donnellan's own view to be, most reactions to his contrast can be classified

according to whether they treat the difference as semantic or as pragmatic. Some treat his contrast as a difference in use alone, while others claim the difference in use is a symptom of a genuinely semantic ambiguity. Pragmatic approaches are now slightly more popular, but the only genuinely uncontroversial claim to be made in this area is that reflection on Donnellan's contrast has forced everyone to be clearer than they were about what the semantics/pragmatics distinction actually is (see **Semantics–Pragmatics Boundary**). In what follows I will outline two paradigmatic treatments, one from each camp.

Pragmatic Treatments (Kripke)

Kripke's role as a critic of Russell's theory of reference is well known (Kripke, 1980), so it might seem surprising that in this context he is Russell's defender. The explanation is simple: Russell's theory of descriptions has two components. On the one hand, and as detailed above, he claimed that definite descriptions have a complex logical form that is not easily read off from their surface structure. But he also claimed that ordinary names are in fact disguised definite descriptions (**Figure 1**). After all, the same puzzles that his theory of the logical form of descriptions is supposed to solve arise equally for ordinary names, so this is a natural extension. In *Naming and necessity*, Kripke rejects the claim that ordinary names are covert definite descriptions but leaves Russell's treatment of definite descriptions as such largely unchallenged. In Kripke (1977) he defends Russell, against Donnellan's criticism at least.

Kripke's response to Donnellan is methodologically elegant (1977: 3c). He imagines a population that has a language like English but, stipulatively, it has a Russellian semantics for definite descriptions. He then asks whether the contrast between referential

and attributive uses of definite descriptions would arise in such a population. It would, he argues. Indeed, Kripke suggests, if speakers of this language did not use 'The F is G' at all, but instead used 'there is an F, identical with anything that is F, and it is G,' an analogue of Donnellan's distinction would still arise. They would use this sentence when they wished to attribute G-ness to something they, mistakenly, believed to be the unique possessor of F-ness. Kripke infers from this that existence of Donnellan's contrast *among speakers of English* does not undermine the claim that English definite descriptions have the semantics Russell says they do.

Kripke's position can be explicated using a famous example due to Paul Grice (1975: 33). A letter of reference for an academic position contains the following evaluation of a weak candidate:

Mr X's command of English is excellent, and his attendance at tutorials has been regular. Yours, etc.

The referee is damning the candidate with faint praise. But in other contexts the sentence could be used without any negative insinuation. This possibility hardly shows that the sentence itself is semantically ambiguous, having a faint-praise sense and a neutral sense. Likewise, Donnellan's contrast does not, according to Kripke, show that S itself is semantically ambiguous. (Grice's own views on Russell's theory are in his 1989 title; see also Neale, 1990: 78; section 3.5.)

The challenge with all pragmatic theories is to provide more than a hand-waving demonstration of how the relevant phenomenon arises systematically from a feature of the conversational context rather than a feature of syntax. Kripke's derivation is based on a distinction he introduces between speaker's referent and semantic referent, which is itself a special case of Grice's distinction between what a speaker means 'on an occasion,' a matter of their intention, and what the expression means 'timelessly.' Roughly, the identity of the semantic referent for a particular use of a referring expression is governed by the conventional rules of the speaker's language or idiolect. These semantic rules establish a condition that something must meet in order to be the referent of the expression whenever it is used. The speaker's referent is what she wishes to talk about and which, so she will typically believe, meets the conditions for being the semantic referent of the expression she therefore uses in an effort to satisfy this wish. Much of the time semantic referent and speaker's referent will be the same entity. But if for example the speaker is mistaken in believing, of the object she wishes to talk about, that it meets the condition for being the semantic referent of the expression she uses, then the two can be distinct.

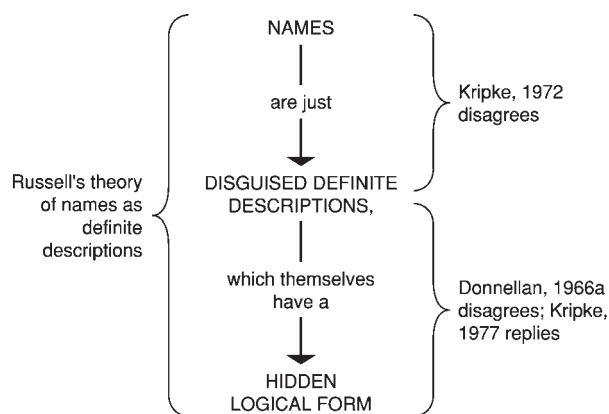


Figure 1 Russell's theory of names as definite descriptions.

Using this distinction, Kripke explains how Donnellan's phenomenon would arise in a language stipulated to assign Russellian truth conditions to definite descriptions. The contrast would emerge out of a distinction between two kinds of wishes its speakers could have when using a definite description. One kind of wish would be to talk about the semantic referent as such, no matter which object that should turn out to be. Another kind of wish would be to say something about a specific object, where use of the definite description would fix the audience's attention on that object whether or not the object is in fact the semantic referent. Kripke claims that this difference in wish would not affect the identity of the semantic referent.

These considerations do not show that English is a Russellian language, with 'the F is G' suffering from no lexical or structural ambiguity. They are compatible with its being the case that some utterances of this sentence are true just in case there is a unique F and it is G, while others are true just in case *an entity the speaker has in mind is uniquely F* is G. But in the absence of independent evidence, Kripke thinks the default assumption should be the simpler one: definite descriptions have a unitary, Russellian semantics. In support of this claim that the no-ambiguity position should be the default, he notes that there is something slightly awkward about describing $U(S)_{REF}$ as true ... about describing the utterance of $U(S)_{REF}$ as strictly true, even if it manages somehow to convey a true proposition.

Semantic Treatments (Wettstein)

Kripke in effect throws down a challenge: show that the difference in use has semantic significance. Wettstein (1981) takes up that challenge. However, his understanding of the attributive/referential distinction differs importantly from Donnellan's. In Donnellan's original examples of referential use, the object referred to fails to be a unique satisfier (in the context) of the descriptive element of the definite description. If all examples were like this, and the existence of a semantic distinction depended on the intuition that such examples were true, the semantic ambiguity thesis would be in trouble. According to Wettstein, Donnellan could and should have illustrated the referential use using only examples in which the object the speaker intends to refer to is a unique satisfier (in the context) of the descriptive element of the definite description. So suppose now that Jones *is* guilty, and no one else is. Wettstein takes the distinction between $U(S)_{REF}$ and $U(S)_{ATT}$ to consist in a difference in the proposition the sentence semantically expresses on each occasion. The proposition expressed by sentences containing a definite

description can have either of two structures, he holds. In referential uses, the proposition is singular. That is, among its constituents, and corresponding to the definite description, is a single object – Jones, say. In attributive uses, the proposition has a general or quantificational structure so that Jones is not a constituent of the proposition. These two propositions are distinct even though the single object must be the unique satisfier of the description for the singular proposition to be true. So Wettstein sees no need to follow Donnellan (as he is often interpreted) in making the further claim that referential uses can be correct even if the referred to object is not the unique satisfier of the description. This further claim is far more tenuous (Kripke, 1977; Lockwood, 1975; Wiggins, 1975).

This clarification of the distinction undermines a widespread reason for denying there is any semantic ambiguity, but it does not show that there is one. Wettstein argues for the positive thesis by noting that referentially used descriptions express a determinately true or false proposition, despite being 'incomplete.' Thus it is possible to utter (8) truly despite there being more than one table in existence.

- (8) 'The table is covered in books'

How is determinacy achieved in the face of incompleteness? According to Wettstein, standard Russellian answers fail. A typical answer would be to say that the incomplete description is tacitly completed in some way. For example, 'table' in (8) is elliptical for (9).

- (9) 'table in the north-east corner of the kitchen at
36 Richmond Way in Harwich, England'

Wettstein responds that how the description is completed is itself indeterminate. It could equally have been (10).

- (10) 'table beside the kitchen window at
36 Richmond Way in Harwich, England'

The same table satisfies each description uniquely, but the descriptions are nonetheless distinct. So no general proposition is uniquely determined. The determinacy of incomplete referential uses *can* be explaining by supposing that definite descriptions function, like demonstratives, to pick out some contextually and conversationally salient table, Wettstein adds. But explaining the determinacy this way has too high a cost for Russellians: the proposition semantically expressed would need to be singular.

Developments

Not everyone has found this argument convincing (see Soames, 1986, Salmon, 1982, Wettstein, 1983, and Reimer, 1998, for discussion). But there are other kinds of argument in favour of interpreting the

contrast as semantic. In a different paper Donnellan himself argued for the ambiguity thesis by considering the behavior of descriptions inside modal operators:

Consider the ambiguity of the counterfactual sentence: "If the President were the author of 'On Denoting,' we would not be in Vietnam." This may be construed as a sentence about Johnson, but there is a second reading on which it is not about him. If it is construed as being about him, then the relevant identity statements give rise to legitimate substitutions; otherwise not. On either reading, the sentence as a whole is non-truth functional, because counterfactual. The ambiguity is not to be explained by supposing that 'the President' refers to Johnson in both cases but that in one case the sentence ascribes a property to him but on the other reading it does not. Rather, on one reading 'the President' does not refer to Johnson at all (Donnellan, 1966b: 687)

Important to the resolution of the issue is the observation, ignored here for brevity, that the referential/attributive contrast may apply to other expressions, including indefinite descriptions, demonstratives, and perhaps even pronouns and names. (For further arguments favouring the semantic position, see Devitt, 1981: 2.5–2.7, 1998, 2004, and Devitt and Sterelny, 1999.)

See also: Descriptions, Definite and Indefinite: Philosophical Aspects; Dthat; Proper Names: Philosophical Aspects; Semantics–Pragmatics Boundary.

Bibliography

- Devitt M (1981). *Designation*. New York: Columbia University Press.
- Devitt M (1998). 'Reference.' In Craig E (ed.) *Routledge encyclopedia of philosophy*. London: Routledge.
- Devitt M (2004). 'The case for referential descriptions.' In Reimer M & Bezuidenhout A (eds.) *Descriptions and Beyond*. Oxford: Clarendon Press.
- Devitt M & Sterelny K (1999). *Language and reality* (2nd edn.). Oxford: Blackwell.
- Donnellan K S (1966a). 'Reference and Definite Descriptions.' *Philosophical Review* 75(3), 281–304.
- Donnellan K S (1966b). 'Substitution and Reference.' *Journal of Philosophy* 63, 685–687.
- Grice H P (1975). 'Logic and conversation.' In Cole P & Morgan J (eds.) *Syntax and Semantics*, vol. 3. London: Academic Press. Reprinted in Grice H P, *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 1989.
- Grice H P (1989). 'Presupposition and conversational implicature.' In Grice H P (ed.) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Kripke S (1972). 'Naming and Necessity.' In Davidson D & Harman G (eds.) *Semantics of Natural Language*. Dordrecht: Reidel. 253–355. Later published as a monograph. Cambridge: Harvard University Press, 1980.
- Kripke S (1977). 'Speaker's Reference and Semantic Reference.' *Midwest Studies in Philosophy* 2, 255–276.
- Kripke S A (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lockwood M (1975). 'On Predicating Proper Names.' *Philosophical Review* 84, 471–498.
- Neale S (1990). *Descriptions*. Cambridge: MIT Press.
- Reimer M (1998). 'The Wettstein/Salmon Debate: Critique and Resolution.' *Pacific Philosophical Quarterly* 79(2), 130–151.
- Russell B (1905). 'On Denoting.' *Mind* 14, 479–493.
- Russell B (1919). 'Descriptions.' In Russell B (ed.) *Introduction to Mathematical Philosophy*. London: George Allen and Unwin. Chapter 16, pp. 167–180. Reprinted in Ostertag G (ed.) *Definite Descriptions: A Reader*. Cambridge, MA: MIT Press.
- Salmon N U (1982). 'Assertion and Incomplete Definite Descriptions.' *Philosophical Studies* 42, 37–46.
- Soames S (1986). 'Incomplete Definite Descriptions.' *Notre Dame Journal of Formal Logic* 27, 349–375. Reprinted in Ostertag G (ed.) *Definite Descriptions*. Cambridge: MIT Press, 275–308, 1998.
- Wettstein H K (1981). 'Demonstrative Reference and Definite Descriptions.' *Philosophical Studies* 40, 241–258. Reprinted in Ostertag G (ed.) *Definite Descriptions*. Cambridge: MIT Press, 1998.
- Wettstein H K (1983). 'The Semantic Significance of the Referential-Attributive Distinction.' *Philosophical Studies* 44, 187–196.
- Wiggins D (1975). 'Identity, Designation, Essentialism and Physicalism.' *Philosophia* 5, 1–30.

Relevance Theory

F Yus, University of Alicante, Alicante, Spain

© 2006 Elsevier Ltd. All rights reserved.

Relevance theory (henceforth RT), a cognitive theory of human communication developed by D. Sperber and D. Wilson, was fully described in their 1986 book (Sperber and Wilson, 1986, 2nd edn., 1995), but it really

emerged in the late 1970s and early 1980s as a cognition-centered alternative to Grice's cooperation-ruled explanation of human communication (see Wilson and Sperber, 1981). Since then, it has been a highly influential theory in pragmatics producing a good number of studies backing it up, criticizing it, or applying it to different pragmatic research areas (see Yus, 1998; Wilson and Sperber, 2002a; and RT Bibliography).

The main assumption of the theory is that human beings are endowed with a biologically rooted ability to maximize the relevance of incoming stimuli (including linguistic utterances and other communicative behavior). Relevance is not only a typical property of external stimuli (e.g., utterances), but also of internal representations and thoughts, all of which may become inputs for cognitive processing. The pursuit of relevance is a typical aspect of the mental activity of human beings, always geared to obtaining the highest reward from the stimuli that they process. This biological endowment is the result of the evolution of the architecture and complexity of the human mind and part of a general human ability to meta-represent one's and other people's thoughts and intentions: "As a result of constant selection pressure towards increasing efficiency, the human cognitive system has developed in such a way that our perceptual mechanisms tend automatically to pick out potentially relevant stimuli, our memory retrieval mechanisms tend automatically to activate potentially relevant assumptions, and our inferential mechanisms tend spontaneously to process them in the most productive way" (Wilson and Sperber, 2002a: 254). Together with another uniquely human cognitive endowment, the ability to metarepresent one's and other people's thoughts and intentions, this tendency to maximize relevance allows us to predict what information is likely to be relevant to other people and what interpretive steps might be involved in its processing, and therefore allows for the manipulation of other people's thoughts.

Basic Claims

Four statements can summarize this theory (Wilson, 1994: 44): (a) the decoded meaning of the sentence is compatible with a number of different interpretations in the same context; (b) these interpretations are graded in terms of accessibility; (c) hearers rely on a powerful criterion when selecting the most appropriate interpretation; and (d) this criterion makes it possible to select one interpretation among the range of possible interpretations, to the extent that when a first interpretation is considered a candidate matching the intended interpretation, the hearer will stop at this point.

These statements can be broken down into a number of basic claims, as summarized below.

Code versus Inference

Unlike the so-called **code model** of communication, according to which messages are simply coded and decoded, Sperber and Wilson favor an **inferential model** in which decoding plays a minor role compared with the inferential activity of the interpreter.

Within this approach, the decoding of utterances underdetermines their interpretation and serves rather as a piece of evidence about the speaker's meaning. Verbal communication does involve the use of a code (i.e., the grammar of the language), but inference plays a major role in turning the schematic coded input into fully propositional interpretations.

One of the most interesting contributions of RT is, precisely, the claim that there is a wide gap between the (coded) sentence meaning and the (inferred) speaker's meaning, which has to be filled inferentially. Comprehension starts at the context-free identification of the utterance's logical form, which is then enriched to yield explicit information (**explicatures**) and/or implicit information (**implicatures**) (*see Implicature*).

A Post-Gricean Theory

Sperber and Wilson acknowledge the filiation of RT from Grice's view of communication, but there are several aspects in which they depart from Grice. This is the reason why we can call RT a post-Gricean theory, a theory that takes the Gricean approach to communication as a mere starting point, as opposed to neo-Gricean theories which stay much closer to Grice's cooperative principle and its maxims (*see Cooperative Principle; Maxims and Flouting*). Several points deserve explanation:

1. One of the major contributions by Grice was to underline the role that intentions (roughly defined as mental representations of a desired state of affairs) play in human communication. His emphasis on the expression and recognition of intentions laid the foundations of the inferential model of communication. Crucially to Grice, the hearer explains the speaker's communicative behavior by identifying the underlying intention, a typically human form of mind-reading activity. However, Sperber and Wilson do not agree with the complex schema of human reasoning that Grice proposed for the derivation of implicatures.

Sperber and Wilson also point out that Grice's emphasis on the role of intentions corroborates the fact that communication can exist without the need for a code. All that the communicator has to do to communicate a thought is to get the addressee to recognize his/her intention to convey it.

Sperber and Wilson distinguish two levels of intention: **informative** (an intention to inform the hearer of something) and **communicative** (the intention to inform the addressee of that informative intention). In inferential communication, the identification of the informative intention is done through the identification of the communicative intention, the process being activated by verbal ostensive

communication, in which it is clear to both speaker and hearer (**mutually manifest** in Sperber and Wilson's terminology) that the speaker has the (metarepresentational) intention to communicate something. Unlike other forms of information transmission, 'ostensive inferential communication' involves both types of intention, and is achieved by ostensively providing an addressee with evidence that helps him/her infer the speaker's meaning.

2. RT explains the hearer's inference of the (intended) speaker's meaning from the coded sentence meaning by resorting to another central claim suggested by Grice: that ostensively communicated utterances automatically generate expectations that activate the hearer's search for the speaker's meaning. But whereas Grice explained these expectations in terms of the assumption by hearers that speakers were observing the cooperative principle and its maxims, within RT these expectations are explained in cognitive terms (basically proposing the existence of a Cognitive Principle of Relevance), without reliance on a cooperative principle.
3. For Sperber and Wilson, no maxims, in the Gricean sense, are required for the explanation of communication. This is especially evident in the case of the Maxim of Quality (roughly, 'tell the truth'), which Grice proposed for the explanation of figurative language and irony. Sperber and Wilson have shown that people are normally 'loose' when they speak and only on very specific occasions do they intend their utterances to be regarded as literally true. In addition, Sperber and Wilson propose that all uses of language, whether loose (metaphor, hyperbole, etc.) literal can be addressed with a single explanatory framework based on general expectations of relevance.

Two Principles of Relevance

Initially, Sperber and Wilson proposed one Principle of Relevance to account for the fact that an act of 'ostension' carries a guarantee of its eventual relevance, but in the Postface to the second edition of their book (Sperber and Wilson, 1995: 260ff.), they propose that we can distinguish a broad **cognitive principle of relevance**: "human cognition tends to be geared to the maximisation of relevance," as well as a narrower (**communicative principle of relevance**: "every act of ostensive communication communicates a presumption of its own optimal relevance"; 1986: 158), the latter being the main focus of analysis within pragmatics. But the former is important, too, because it stresses the fact that we are biologically geared toward processing the most relevant inputs available. In addition, it is this evolved disposition

that allows for the prediction of the mental states of others, which is crucial in human communication.

The communicative principle involves a definition of optimal relevance comprising two parts: (a) The ostensive stimulus is relevant enough for it to be worth the addressee's effort to process it; and (b) The ostensive stimulus is the most relevant one compatible with the communicator's abilities and preferences (Sperber and Wilson, 1995: 267, 270). As Wilson and Sperber (2002a: 257–258) correctly point out, communicators "cannot be expected to go against their own interests and preferences in producing an utterance. There may be relevant information that they are unable or unwilling to provide, and ostensive stimuli that would convey their intentions more economically, but that they are unwilling to produce, or unable to think of at the time." All this is covered by clause (b) of the definition of optimal relevance, which states that the ostensive stimulus is the most relevant one "that the communicator is **WILLING AND ABLE** to produce" (Sperber and Wilson, 1995: 258).

Assessing Relevance: Cognitive Effects versus Processing Effort

Unlike what is the case in 'static' pragmatics, which foregrounds the importance of context but somehow takes it for granted or is merely interested in dissecting, as it were, its elements, Sperber and Wilson's theory views the context as a dynamic, mental entity made up of a subset of the person's assumptions about the world; it is this subset that is accessed in the search for relevance. Often several extensions of context are required to arrive at an optimally relevant interpretation, but as soon as one interpretation is found to be satisfactory, interpretation stops and no other interpretive hypotheses are considered: "When a hearer following the path of least effort finds an interpretation which satisfies his expectations of relevance, in the absence of contrary evidence, this is the best possible interpretive hypothesis" (Wilson and Sperber, 2002b: 605).

The aforementioned Communicative Principle of Relevance predicts a basic procedure for hearers when hypothesizing about contextual extensions required for the interpretation of a verbal stimulus: to consider interpretive hypotheses in order of accessibility (following a path of least effort) and to stop when they arrive at an interpretation which satisfies the expectations of relevance raised by the stimulus itself. Relevance, then, is a matter of balance between the interest that the utterance might provide (in terms of so-called 'positive cognitive effects') and the mental effort that obtaining this interest demands.

Relevance is a characteristic of an input to the human cognitive processes which, when processed in a certain

context, yields positive cognitive effects. Because there are too many possible stimuli to which we can pay attention, our cognitive architecture is designed to allocate our processing effort in such a way that benefit is maximized. Hence, relevance has to do with the improvement of the person's knowledge; this can be achieved either by adding new information, by revising existing assumptions, or by yielding new conclusions resulting from the combination of old and new information (in this case **contextual implicatures** are generated).

The definition of relevance of an input to an individual involves two clauses: "(a) everything else being equal, the greater the positive cognitive effects achieved in an individual by processing an input at a given time, the greater the relevance of the input to that individual at that time; and (b) everything else being equal, the smaller the processing effort expended by the individual in achieving those effects, the greater the relevance of the input to that individual at that time" (Wilson and Sperber, 2002b: 602).

Current Issues and Open Debates

The Explicit/Implicit Distinction

One of the key differences between Grice's model and Sperber and Wilson's lies in the demarcation of explicit and implicit communication. For Grice, **what is said** involved little inference, mainly reduced to disambiguation and reference assignment, while all the inferential load was laid upon the derivation of implicatures, the latter being obtained **after** an interpretation reduced to the literal meaning has been found inappropriate, in a so-called **dual-stage processing**. Sperber and Wilson reject this view and favor a more adequate, **mutual parallel adjustment** of explicit content – explicatures – and implicit import – implicatures – during interpretation, and with no preconceived sequential arrangement.

Within RT, explicitly communicated information not only demands as much contextualization as do implicatures, but also covers aspects of communicated meaning which Grice included in the term implicature (e.g., the so-called **generalized conversational implicatures**, most of which are now pictured as explicit information, see Levinson, 2000; Carston, 2002).

In addition to implicatures, Sperber and Wilson propose two types of explicitly communicated information: the basic-level **explicature**, and the **higher-level explicature**. The latter also includes the speaker's attitude (*to regret that ... to be happy that ... etc.*) or a higher-order speech-act schema (*to be asking that ... to be ordering that ... etc.*). Both explicatures and implicatures allow for degrees (i.e., strong and weak

explicatures/implicatures), depending on the addressee's responsibility for their derivation and the amount of mental processing required.

Other notions used by other authors in the definition of explicit information, for instance **literal meaning** or **what is said**, are put into question by Sperber and Wilson, because these do not play any useful role in the study of verbal comprehension: "even when a literal meaning is available, it is not automatically the preferred interpretation of an utterance. In fact, literalness plays no role in our account of language comprehension, and nor does the notion of what is said" (Wilson and Sperber, 2002b: 586). This is because, among other reasons, hearers commonly derive loose interpretations rather than purely literal ones: "hearers have no objection to strictly false approximations as long as the conclusions they bother to derive from them are true. In fact, they might prefer the shorter approximations to their longer-winded but strictly true counterparts for reasons of economy of effort" (Wilson and Sperber, 2002b: 598).

Whereas Bach (1994) has proposed a third term, **implicature**, half-way between explicatures and implicatures, other authors, such as Vicente (2002), reject this blurring of the explicit/implicit dichotomy. The term 'implicature' covers several cases which would fit into Sperber and Wilson's notion of explicature, basically being completions of the semantic representation of the sentence (e.g., 'The table is too wide' [to go through the door]) and nonliteral uses of sentences in which no constituent is being used non-literally, what Bach calls **standardized nonliterality** (e.g., (said to a person who has cut himself) 'You are not going to die' [from this cut]).

Conceptual and Procedural Encoding

One of the most interesting lines of research within relevance theory is the one that differentiates between **conceptual meaning** and **procedural meaning**. Wilson and Sperber (1993: 10) summarize this dichotomy as follows: "inferential comprehension involves the construction and manipulation of conceptual representations; linguistic decoding feeds inferential comprehension; linguistic constructions might therefore be expected to encode two basic types of information: concepts or conceptual representations on the one hand, and procedures for manipulating them on the other."

Most words encode concepts, but some words give instructions as to how conceptual representations are to be manipulated and hence encode procedural meaning. Blakemore and her followers applied the notion to connectives (Blakemore, 1987) and discourse markers (Blakemore, 2002), which constrain the inferential phase by indicating the kind of

inferential process that the hearer should go through (hence reducing the eventual overall effort) in the subsequent stretch of discourse. In recent research, the list of procedural items has been extended to cover nonverbal elements such as intonation (*see Meaning: Procedural and Conceptual*).

***Ad hoc* Concept Formation**

The notion of *ad hoc* concept construction is one of the latest developments of relevance theory in the area of figurative language (especially metaphors), which has also been extended to the analysis of how concepts in general are processed (cf. Carston, 2002; Pilkington, 2000).

The traditional relevance-theoretic account of figurative language relies on the assumption that there is an interpretive resemblance between the coded concept and the intended thought. And we can say the same about the whole utterance whose propositional form resembles the propositional form of the communicator's thought (Pilkington, 2000: 90). From this viewpoint, the interpretive resemblance between, for instance, a coded metaphor and the thought which it resembles would lead to the hearer's derivation of stronger/weaker implicatures.

Within an alternative account of utterance interpretation, it is claimed that the metaphor provides a new *ad hoc* concept for the proposition expressed by the utterance (instead of favoring the derivation of implicatures) (*see Metaphor: Psychological Aspects*). Encyclopedic entries would be explored in such a way that an increase in the *salience* of a number of assumptions is created, providing an encyclopedic entry for the new concept (Pilkington, 2000: 95-96; for 'salience,' see Giora, 2002). They are *ad hoc* "because they are not linguistically given, but are constructed online in response to specific expectations of relevance raised in specific contexts. There is a difference then between *ad hoc* concepts, accessed by a spontaneous process of pragmatic inference, and lexicalized concepts, which are context-invariant" (Carston, 2002: 322).

Mutual Knowledge versus Mutual Manifestness

Sperber and Wilson reject the traditional notion of mutual knowledge because it generates an endless recursion (A knows that p, B knows that A knows that p, A knows that B knows that A knows that p, and so on). Instead, they propose the notion of **mutual manifestness** (*see Sperber and Wilson, 1990*). What is 'manifest' is what one is capable of inferring or capable of perceiving, even if one hasn't done so yet. The sum of all the manifest assumptions is the person's **cognitive environment**. A set of assumptions manifest to several individuals constitutes their

shared cognitive environment. When it is manifest to all the people sharing a cognitive environment that they share it, then, this is a **mutual cognitive environment**, made up of mutually manifest assumptions. Communication is a matter of making certain assumptions mutually manifest to both speaker and hearer.

Several authors have criticized the notion of mutual manifestness. For example, Mey and Talbot (1988) point out that what Sperber and Wilson do is to send mutual knowledge out at the front door and then let it in at the back, disguised as 'mutually manifest assumptions.' For these authors, cognitive environments are not distinguishable from mutual knowledge; thus, Sperber and Wilson appear to be using the same concept that they want to abandon. To my knowledge, neither Sperber and Wilson nor their critics have been persuaded to abandon their differing claims on mutuality.

Communicated and Noncommunicated Acts

One of the most underdeveloped areas within relevance-theoretic research is the relationship between RT and speech acts. In short, Sperber and Wilson (1986: 244-246) distinguish between communicated and noncommunicated acts. The former depend on the addressee's perception that a certain speech act has been performed (e.g., admitting, promising, thanking), while in non-communicated acts, communication does not depend on the identification of a particular speech act (e.g., predicting, warning, suggesting). In this case, successful communication lies in the hearer's recovery of adequate cognitive effects from the utterance with the aid of context and in the recovery of the speaker's intentions.

In a recent paper, Nicolle (2000) has argued against the existence of noncommunicated speech acts in the RT sense. Examples of noncommunicated acts (such as the act of warning in "The path is slippery here") are reconsidered by Nicolle in social terms, and their influence on the interlocutors' social environments implies that they also have to be communicated: "the recovery of information relating to social relations is an essential element of the comprehension process. When the recovery of such information depends on the identification of a particular speech act, that speech act is by definition a communicated act" (Nicolle, 2000: 239).

Irony and the Notion of Echo

In Wilson and Sperber (1992), the authors conceptualize irony in **interpretive** terms. An ironic utterance is an interpretation of another thought, utterance, or assumption that it resembles and which the speaker attributes to a different speaker or to himself/herself

at another time. Ironic utterances are echoic, that is, they simultaneously refer to an attributed thought – or utterance, or assumption – and express an attitude to it. More specifically, the speaker's attitude toward what is echoed has to be dissociative. This dissociation may apply to either the proposition expressed by the utterance, or to some effect that is generated by that utterance.

Several authors have commented upon this proposal. For instance, in some of the papers collected in Carston and Uchida (1998) it is claimed that irony can be nonechoic. In their reply, Sperber and Wilson (1998) maintain that although most utterances cannot be understood as echoic (i.e., there is no accessible representation that they might be taken to echo), an utterance has to be echoic to be interpreted as ironical (see *Irony*).

Modularity

Initially, Sperber and Wilson adopted the view of the mental architecture of the mind proposed by Jerry Fodor in the early 1980s: several modules feeding a central processor with specific information.

Modules are evolved, special-purpose mental mechanisms, typically automatic, informationally encapsulated, and domain-specific. For instance, the language module is only (and automatically) activated by verbal stimuli, feeding the central processor with a schematic logical form which then has to be enriched inferentially.

Over the last few years, this view of the mind has changed within RT (and also within evolutionary psychology), especially concerning the structure of the central processor, which is also regarded to be modular (Carston, 1997; Sperber and Wilson, 2002; Wilson, 2003). The most important module, specifically a submodule of the general 'theory of mind' ability, is **the pragmatic module**, which also exhibits qualities typically associated with modules. For example, this pragmatic module is biologically endowed, only activated by a specific type of information (ostensively communicated information), and constrained by its own principle: the Communicative Principle of Relevance.

Relevance Theory as Asocial

RT has been criticized for being hyperindividualistic and for avoiding the social aspects of communication (Mey and Talbot, 1988). Sperber and Wilson (1997: 147) acknowledge that they have concentrated on the inferential activity of the individual, but inferential communication is **also** essentially social: "Inferential communication is intrinsically social, not just because it is a form of interaction, but also, less trivially, because it exploits and enlarges the scope of basic

forms of social cognition. Right or wrong, this is a strong sociological claim." Although Sperber and Wilson have not studied uses of communication to convey information about the social relationship between the interlocutors, they do not mean to deny its importance, or to express a lack of interest in the issues or the work done; they merely feel that, at this stage, they can best contribute to the study of human communication by taking it at its most elementary level, and abstracting away from these more complex (socially connoted) aspects (Sperber and Wilson, 1997). Hence, for them, although "so far, the contribution of relevance theory to the study of human communication has been at a fairly abstract level... it seems to us to have potential implications at a more concrete sociolinguistic level" (Sperber and Wilson, 1997: 148).

A proposal by Escandell-Vidal (2004) aims at integrating inferential and social issues in terms of principles and norms, respectively, and as part of a domain-specific picture of mental activity. The mind operates according to principles that are in charge of obtaining fully propositional interpretations from coded stimuli. When dealing with norms, the mind is engaged in both a long-term and a short-term task. The short-term one analyzes and categorizes incoming information, and the long-term task builds up and updates socially accepted behavior.

Empirical Evidence

A common criticism of RT is that it is highly speculative, predicting without empirical evidence the mental procedures and interpretive steps the human mind goes through in human communication. Obviously, we are dealing with an object of study, the human mind, which is highly complex and still largely unexplained.

Sperber and Wilson (2002: 143) acknowledge that in much pragmatic research, there is a certain reluctance to get down to experimentation. However, relevance theorists have been particularly eager to combine theoretical issues with all the possibilities of testing provided by the careful use of linguistic intuitions, observational data, and the experimental methods of cognitive psychology (see Wilson and Sperber, 2002b: 607, note 7 for references). Recent research has aimed at an empirical explanation of central claims of the theory. For instance, Van der Henst and Sperber (2004) review various experimental tests of the two Principles of Relevance. They claim that the hypothesis that hearers spontaneously rely on a relevance-guided interpretive procedure can be experimentally tested either by intentionally manipulating the effort required to process a stimulus

or by changing the order of accessibility of several competing interpretations for the same stimulus. Another possible test is a manipulation of the effect factor by making a specific interpretation more or less likely to satisfy the expectations of relevance.

Other studies have focused on other possible areas of RT-based empirical research. Among them we can underline the ones on the Wason selection task. For example, Sperber *et al.* (1995) tested how participants derive implications from conditional statements in order of accessibility, stop when their search for relevance reaches an adequate balance of cognitive effects and processing effort, and select cards on the basis of this interpretation. The authors were able to manipulate the effort and effect factors by varying the content and context of the conditional statement, so as to elicit correct or incorrect selections at will (cf. Wilson and Sperber, 2002a: 279).

The plausibility of a Gricean maxim of truthfulness to explain human communication has also been tested (Van der Henst *et al.*, 2002). These authors showed that when people ask a stranger the time in the street they get, as a reply, “a time that is either accurate to the minute or rounded to the nearest multiple of five, depending on how useful in the circumstances they think a more accurate answer would be” (Wilson and Sperber, 2002b: 598), regardless of whether the people asked have (accurate) digital watches. These rounded answers are not strictly true, but they are easier for their audience to process.

Applications

RT has been applied to a number of research areas, among which we can distinguish the following.

1. **Grammar.** For RT the interest lies in how grammatical attributes constrain the choice of a candidate interpretation. In this sense, the grammatical arrangement of utterances plays an important part throughout this cognitive contextualization. From this point of view, several aspects of grammar have been addressed, including connectives (often within a conceptual/procedural account), conditionals, modals and modality, adverbs and adverbials, mood(s), tense(s), the article, etc.
2. **Humor.** Within a relevance-theoretic approach, humor is no longer a property of texts and, instead, what we need to characterize are the audience's mental processes in the interpretation of humorous texts.

Underlying this approach to humor lies the premise that communicators can predict and manipulate the mental states of others. Knowing that the addressee is likely to pick out the most relevant

interpretation of the joke (or some part of it), the humorist may be able to produce a text that is likely to lead to the selection of an accessible interpretation, which is then invalidated at some point. In Yus (2003), for instance, it is claimed that in many jokes the initial part has multiple interpretations, which are graded according to their accessibility. The hearer is led to select an overt (i.e., relevant) interpretation of this part of the joke. Suddenly, the hearer notices that the subsequent part has a single covert interpretation which is eventually found to be the correct one (and the one providing a coherent interpretation to the whole text) and which humorously surprises the hearer.

3. **Media discourses.** RT has also been successfully applied to the interpretation of media discourses, including films, newspaper headlines, comics, Internet discourse, and advertising. The last one is probably one of the most extensive applications of the theory. The control over the amount of information provided, the predictability of consumers' responses, and the calculation of the effort required to process information, all typical features of the strategies by the advertisement makers, can easily be analyzed using a relevance-theoretical approach (see Tanaka, 1994).
4. **Literature.** Several studies have applied RT to literary discourse (Pilkington, 2000 is an example). Within an RT approach, literariness has to be analyzed as cognitive effects triggered by textual stimuli, involving special mental processes which, through a relevance-driven cognitive exploration, result in the marginally increased salience of a wide range of assumptions (Pilkington, 2000: 189). Because in literature it is more difficult (if not impossible) to make assumptions mutually manifest, a greater load of responsibility is laid upon the reader in extracting the intended (or, alternatively, his/her own) interpretation of the text, plus whatever feelings and emotions are associated with the comprehension of the text.
5. **Politeness.** This is a typical social feature of communication that somehow appears not to suit the individual-centered approach within RT. However, several studies have attempted an explanation of politeness in relevance-theoretic terms. For instance, politeness has been explained within RT as a verbal strategy compatible or incompatible with the background expectations about the current relationship holding between speaker and hearer, thus leading to different relevance-oriented interpretive paths (see Jary, 1998).
6. **Translation.** In RT-bases studies such as Gutt's (2000), there is a tendency to exploit the idea of

resemblance between the intended interpretations of utterances. (For an account of relevance-theoretic applications to translation.)

Concluding Remarks

In RT, Sperber and Wilson propose a coherent cognitive account of how the human mind proceeds when attempting to select a plausible interpretation of ostensibly communicated stimuli. They rely on the hypothesis that a biologically rooted search for relevance aids human beings in the inferential enrichment of typically underdetermined coded texts and utterances resulting in fully propositional interpretations.

The theory has provided insights in several debates in pragmatics and cognitive science and has been applied to a good number of research areas. Undoubtedly, RT will continue to stir fruitful intellectual debates on the explanation of human communication.

See also: Cognitive Science and Philosophy of Language; Cooperative Principle; Implicature; Intention and Semantics; Irony; Maxims and Flouting; Meaning; Procedural and Conceptual; Metaphor: Psychological Aspects.

Bibliography

- Bach K (1994). 'Conversational implicature.' *Mind and Language* 9, 124–162.
- Blakemore D (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore D (2002). *Relevance and linguistic meaning. The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.
- Carston R (1997). 'Relevance-theoretic pragmatics and modularity.' *UCL Working Papers in Linguistics* 9, 29–53.
- Carston R (2002). *Thoughts and utterances*. Oxford: Blackwell.
- Carston R & Uchida S (eds.) (1998). *Relevance theory. Applications and implications*. Amsterdam: John Benjamins.
- Escandell-Vidal V (2004). 'Norms and principles. Putting social and cognitive pragmatics together.' In Márquez-Reiter R & Placencia M E (eds.) *Current trends in the pragmatics of Spanish*. Amsterdam: John Benjamins. 347–371.
- Giora R (2002). *On our mind*. Oxford: Oxford University Press.
- Gutt E-A (2000). *Translation and relevance: cognition and context* (2nd edn.). Manchester: St Jerome Publishing.
- Jary M (1998). 'Relevance theory and the communication of politeness.' *Journal of Pragmatics* 30, 1–19.
- Levinson S C (2000). *Presumptive meanings: the theory of Generalized Conversational Implicatures*. Cambridge, Mass.: MIT Press.
- Mey J L & Talbot M (1988). 'Computation and the soul.' *Semiotica* 72, 291–339.
- Nicolle S (2000). 'Communicated and non-communicated acts in relevance theory.' *Pragmatics* 10, 233–245.
- Pilkington A (2000). *Poetic effects. A relevance theory perspective*. Amsterdam: John Benjamins.
- RT Bibliography. <http://ua.es>. (University of Alicante).
- Sperber D & Wilson D (1986). *Relevance: communication and cognition*. (2nd edn., 1995). Oxford: Blackwell.
- Sperber D & Wilson D (1990). 'Spontaneous deduction and mutual knowledge.' *Behavioral and Brain Sciences* 13, 179–184.
- Sperber D & Wilson D (1997). 'Remarks on relevance theory and the social sciences.' *Multilingua* 16, 145–151.
- Sperber D & Wilson D (1998). 'Irony and relevance: a reply to Seto, Hamamoto and Yamanashi.' In Carston R & Uchida S (eds.), 283–293.
- Sperber D & Wilson D (2002). 'Pragmatics, modularity and mind-reading.' *Mind and Language* 17, 3–23.
- Sperber D, Cara F & Girotto V (1995). 'Relevance theory explains the selection task.' *Cognition* 57, 31–95.
- Tanaka K (1994). *Advertising language. A pragmatic approach to advertisements in Britain and Japan*. London: Routledge.
- Van der Henst J-B, Carles L & Sperber D (2002). 'Truthfulness and relevance in telling the time.' *Mind and Language* 17(5), 457–466.
- Van der Henst J P & Sperber D (2004). 'Testing the cognitive and the communicative principles of relevance.' In Noveck I & Sperber D (eds.) *Experimental Pragmatics*. London: Palgrave.
- Vicente B (2002). 'What pragmatics can tell us about (literal) meaning: a critical note on Bach's theory of implicature.' *Journal of Pragmatics* 34, 403–421.
- Wilson D (1994). 'Relevance and understanding.' In Brown G et al. (eds.) *Language and understanding*. Oxford: Oxford University Press. 37–58.
- Wilson D (2003). 'New directions for research on pragmatics and modularity.' *UCL Working Papers in Linguistics* 15, 105–127.
- Wilson D & Sperber D (1981). 'On Grice's theory of conversation.' In Werth P (ed.) *Conversation and discourse*. London: Croom Helm. 155–178.
- Wilson D & Sperber D (1992). 'On verbal irony.' *Lingua* 87, 53–76.
- Wilson D & Sperber D (1993). 'Linguistic form and relevance.' *Lingua* 90, 1–25.
- Wilson D & Sperber D (2002a). 'Relevance theory.' *UCL Working Papers in Linguistics* 14, 249–290.
- Wilson D & Sperber D (2002b). 'Truthfulness and relevance.' *Mind* 111(443), 583–632.
- Yus F (1998). 'A decade of relevance theory.' *Journal of Pragmatics* 30, 305–345.
- Yus F (2003). 'Humor and the search for relevance.' *Journal of Pragmatics* 35, 1295–1331.

Representation in Language and Mind

F Egan, Rutgers University, New Brunswick, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

Languages are human artifacts used for a variety of purposes, most notably, communication. Mental states, such as beliefs, desires, intentions, hopes, and fears – the so-called ‘propositional attitudes’ – are internal states of organisms that are, typically, both effects of external conditions (and other mental states) and causes of behavior. Languages and states of mind are said to ‘represent’ – they can be about things other than themselves. They can even represent nonexistent states of affairs, such as, for example, Superman’s working at the *Daily Planet*. One might wonder how it is possible for linguistic objects – utterances and inscriptions – and states of mind, both of which are physically realized, to represent, i.e., to have meaning or content. And one might wonder how a particular linguistic object or a particular mental state represents what it does. In virtue of what fact, for example, does a thinker’s mental state mean *it will be raining soon*, rather than $2 + 2 = 4$? This tangle of issues is known as the ‘problem of intentionality’.

The Relationship between Language and Thought

It is plausible that representation in language and representation in thought are related. A speaker’s utterances typically express his or her thoughts. One can both think and say that it will be raining soon. Perhaps the limits of language are the limits of thought. Many theorists have found plausible the hypothesis that we can only think thoughts that can be expressed in public language. In any case, a theory of representation should explain how language and thought are related.

Several general strategies for explaining representation in language and mind are possible:

1. Mental representation might be construed as basic or fundamental, and the representational properties of utterances and inscriptions explained in terms of the mental states (beliefs, intentions, and so on) of speakers, hearers, writers, and readers. We shall call such theories ‘mind-based’ theories.
2. Linguistic meaning might be taken as basic and the meanings of mental states explained in terms of the thinker’s linguistic capacities. We shall call these theories ‘language-based’ theories. They have the apparent consequence that nonhuman animals and prelinguistic humans do not have

representational mental states – in other words, they do not literally have thoughts.

3. Neither thought nor language is construed as basic, though there may be some explanatory relations between aspects of each. We will call theories of this third type ‘non-reductive’ theories.

Mental Representation as Basic

The most popular theories of representation are mind-based, construing mental representation as basic and linguistic meaning as derivative of it. This explanatory strategy follows the program set out by Paul Grice in his seminal 1957 paper, ‘Meaning’. Grice provides an analysis of a speaker’s meaning something by an utterance on a particular occasion in terms of the speaker producing an utterance with a set of complex intentions to produce certain states of mind in hearers. Grice’s analysis takes for granted the propositional attitudes of speakers and hearers. The question, then, is how these mental states acquire their meaning.

Some theorists of mind construe mental representation as itself language-like. Thought, they claim, takes place in an inner language or code (*see Mentalese*). Propositional attitudes, in this view, are relations to internal sentences. But whether or not the language-of-thought hypothesis is true, a theory of mental representation requires an account of how propositional attitudes acquire their meaning. And if the account is to serve as the basis for an explanation of representation in public language, then it must not presuppose public language meaning.

Mind-based theories are typically ‘naturalistic’: they aim to explain the intentionality (meaning, aboutness) of a subject’s mental states in terms of non-semantic and non-intentional properties of those states. The intuition behind the naturalistic constraint is clear. An adequate account of intentionality – one that dissolves the air of ontological strangeness surrounding it – should explain it in terms compatible with the broader conception of nature provided by the natural sciences.

Information-Based Theories

Physical systems carry information about the external conditions that impinge on them. The number of rings in a tree’s trunk reliably varies with the number of growing seasons; hence the tree’s trunk carries information about, or represents, the age of the tree. Exploiting this general phenomenon, ‘information-based’ semantic theories take the meaning of a mental state or structure to be determined by the cause of its

tokening in certain specifiable conditions. For example, a mental structure means *cat* if it is reliably caused by the presence of cats.

Mental states not only represent; they can occasionally **misrepresent**. Information-based theories have trouble accounting for misrepresentation. If a given mental structure is reliably caused by the presence of cats, then it is likely to be (reliably) caused by other things in less than optimal conditions, for example, by small dogs on a dark night. Since small dogs on a dark night are among the reliable causes of the structure, then, according to the theory, it means *cat or small dog on a dark night*. Rather than misrepresenting a small dog on a dark night as a cat, the mental state genuinely represents its disjunctive cause. Misrepresentation does not seem possible.

Proponents of information-based theories have attempted to solve the so-called ‘error problem’ by levying additional conditions on the circumstances in which a mental state’s meaning is determined by the cause of its tokening. (See, for example, Fodor’s 1990 account, which requires for a mental structure X to mean *cat* not only that it be reliably caused by cats, but also that if any non-cats cause X, then non-cat’s causing X are **asymmetrically dependent** upon cats causing X; that is, non-cats would not cause X unless cats did, but not vice versa.) Whether these modifications handle the error problem without raising further difficulties is a matter of some dispute.

Teleological Theories

According to ‘teleological’ semantic theories, the meaning of a mental state or structure is determined by its biological function, i.e., that for which it was selected. (See Millikan, 1984, for the most developed proposal of this sort.) A variant of the error problem arises for theories that attempt to ground representation in biological function. A simple example makes the point. A frog will snap its sticky tongue at any fly moving in its vicinity. Some state of the frog’s visual system represents the fly, but is the meaning of this state *fly*, *food*, or – given that the frog will react identically to any small dark object moving in its visual field – *small dark moving thing*? An appeal to the biological function of the frog’s internal state will not settle the issue, as there seems nothing to choose among describing the function of the state as *detecting flies*, as *detecting food* (given that flies are the only available food), or as *detecting small dark moving things* (given that almost all small dark moving things in the frog’s vicinity are flies and hence food). A similar sort of indeterminacy is likely to plague efforts to establish the meaning of states of more complex systems, such as humans. The notion of

biological function seems too coarse an instrument to determine representational content.

Conceptual Role Theories

‘Conceptual role semantics’ (CRS) is the most popular theory of mental representation. Versions of CRS have been developed and defended by philosophers, cognitive psychologists, and (under the guise of ‘procedural semantics’) computer scientists. (See Block, 1986, for a general defense of CRS.) According to CRS, a mental state means what it does in virtue of its causal role in the subject’s mental life. What determines that a mental structure means *cats are furry* is the structure’s role in perception, deductive and inductive inference, deliberation, action, and so on. Tokenings of the structure tend to be caused by the presence of furry cats and to cause one to conclude that if something is not furry then it is not a cat. The causal role of a structure meaning *dogs are needy* involves appropriate relations to dogs and instantiations of the property *neediness*.

A problem for CRS arises from the fact that a mental state’s meaning is determined by the totality of its connections to perceptual stimuli, behavior, and other mental states. In this view, all of a subject’s beliefs involving a concept contribute to its meaning. However, no two people have exactly the same beliefs involving any concept. If I believe that bats are scary but you do not, then, according to CRS, our ‘bat-concepts’ are different. Consequently, our beliefs involving our respective bat-concepts, even such core beliefs as each of us would express by the words ‘Bats are mammals,’ have different content. A consequence of the holistic determination of meaning is that no two subjects will share any concepts or any beliefs.

A proponent of CRS might try various maneuvers to mitigate the effects of holism. It is surely implausible to count every belief as contributing equally to the meaning of a concept. CRS needs some notion of **semantic importance** to distinguish the beliefs that contribute heavily to the meaning of associated concepts from those that do not. Whether such a notion can be spelled out without appeal to semantic or intentional notions – that is, without violating the naturalistic constraint – is an open question.

Constraints on a Theory of Mental Representation

In addition to accounting for the possibility of misrepresentation, and not leading to an unacceptable holism, a theory of mental representation must satisfy a ‘cogency condition’: the interpretation of a subject’s mental states licensed by the theory must, by and large, make sense. The theory must not license an interpretation that makes the subject wildly

irrational. The status of the cogency constraint is a matter of some controversy. Some philosophers (for example, Davidson (1984)) have thought it ‘constitutive’ of the intentional framework that thinkers are, by and large, rational. Others have thought only that, as a matter of empirical fact (perhaps because they are products of evolutionary adaptation), thinkers are typically rational.

An adequate theory of mental representation must also account for the fact that thought is fine-grained. The theory should explain how a thinker can believe that Cicero was a Roman orator but fail to believe that Tully was a Roman orator, even though “Cicero” and “Tully” are names for the same individual. This is known as ‘Frege’s problem’. Finally, as noted above, a theory should explain how thinkers can have thoughts about nonexistent objects or states of affairs, such as Superman’s working at the *Daily Planet*.

Linguistic Representation as Basic

Norms-Based Theories

Language-based theories typically explain mental representation in terms of social practices governing overt speech (see, for example, Brandom, 1994). In this view, the social practices that confer meaning on utterances involve norms governing the correct use of certain expressions and the appropriate consequences of such uses. A given utterance-type justifies (that is, permits) certain other utterances and excludes certain others. A speaker who utters ‘There is a dog’ is committed to the assertion *There are some dogs* and to the denial of *There are no dogs* and is taken to be so committed by others. The relevant norms govern the actions of speakers and hearers, where action is not to be understood as presupposing intentional thought, but rather as behavior shaped by communal practices of conforming and censuring, what Brandom (1994) calls ‘the game of giving and asking for reasons’. It is the communal practice of subjects’ taking each others’ performances as ‘correct’, ‘incorrect’, ‘entitling’, or ‘excluding’ other performances that confers meaning on these performances, and, derivatively, on the subjects’ states of mind. The norms governing these practices make intentionality of both speech and thought possible.

An account of this sort will succeed only if notions of ‘correctness’, ‘entitlement’, ‘commitment’, and so on, can themselves be understood in non-intentional and non-semantic terms. If we must presuppose that the relevant performances – both the utterances themselves and others’ attitudes toward these utterances – are meaningful, if, for example, we must presuppose

that these performances either involve the use of antecedently meaningful expressions or consist in actions produced with certain intentions in mind, then their meaningfulness does not reside in, or result from, the conforming and censuring practices themselves. The project fails if the social practices that supposedly confer meaning cannot be explained without presupposing that the performances governed by these practices have meaning. It is not obvious that this explanatory burden has been discharged.

A Non-Reductive Proposal

A non-reductive construal of the norms-based project is possible. One might see the task of articulating the structure of the norms governing overt speech as exposing how the intentionality of thought and the (public) norms governing overt speech are intimately connected. Such an account would aim to show that we cannot understand representation, in either language or thought, without a notion of inference, that is, without the idea of certain claims entitling or justifying others. (This idea is a cornerstone of the CRS developed by Wilfred Sellars.) On this construal, the project is an attempt to work out the interconnections between language and thought, without taking either as basic.

See also: Mentalese.

Bibliography

- Block N (1986). ‘Advertisement for a Semantics for Psychology.’ In French P A, Uehling T E Jr & Wettstein H K (eds.) *Midwest studies in philosophy 10: Studies in the Philosophy of Mind*. Minneapolis: University of Minnesota Press.
- Brandom R B (1994). *Making it explicit: reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.
- Cummins R (1989). *Meaning and mental representation*. Cambridge, MA: MIT Press.
- Davidson D (1984). *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
- Fodor J A (1990). ‘A theory of content, I, II.’ In Fodor J A (ed.) *A theory of content and other essays*. Cambridge, MA: MIT Press. 51–136.
- Frege G (1892). ‘On sense and reference.’ Black M (trans.). Reprinted in Ludlow P (ed.). *Readings in the philosophy of language*. Cambridge, MA: MIT Press. 563–583.
- Grice H P (1957). ‘Meaning.’ *Philosophical Review* 66, 377–388.
- Millikan R (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.
- Putnam H (1975). ‘The meaning of ‘meaning’.’ In Putnam H (ed.) *Philosophical papers*, vol. 2. *Mind, language*,

and reality. Cambridge: Cambridge University Press. 215–271.
 Sellars W (1956). ‘Empiricism and the philosophy of mind.’ In Scriven M, Feyerabend P & Maxwell G (eds.)

Minnesota Studies in the Philosophy of Science, vol. 1. Minneapolis: University of Minnesota Press. 253–329.
 Stich S P & Warfield T A (eds.) (1994). *Mental representation: a reader*. Oxford: Blackwell.

Rigid Designation

S Predelli, University of Nottingham, Nottingham, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Rigidity is a semantic notion. Semantics studies the relationships between expressions in a language and (typically) extralinguistic items. Examples of semantically interesting features are the relationship between the name ‘George Bush’ and Bush, the individual to which it refers; the relationship between the sentence ‘George Bush is a liberal’ and falsehood, its truth value; and, at least in some views, the relationship between the definite description ‘the President of the United States’ and Bush, the individual it denotes.

The task of a semantic theory is not that of immediately assigning the aforementioned features to the appropriate expressions. For instance, that the sentence ‘George Bush is a liberal’ is to be assigned falsehood depends not only on the semantic traits of the expressions under analysis, but also on obviously extrasemantic facts having to do with Bush’s political convictions. By the same token, that the *denotatum* of ‘the President of the United States’ is George Bush, rather than John Kerry or myself, also depends on extrasemantic facts having to do with the results of the latest presidential election. For this reason, the aim of a semantic theory is that of presenting *relativized* assignments of semantic features. For instance, what our theory may conclude is that ‘George Bush is a liberal’ is true with respect to all and only those parameters that provide a positive reply to the query about Bush’s leftist tendencies, and that ‘the President of the United States’ denotes Bush with respect to all and only those parameters with respect to which Bush won the presidential elections. According to a customary approach, parameters of this type may be understood as possible worlds. In a more complex framework, the *relata* appropriate for semantic evaluation may include, together with possible worlds, a temporal parameter: for instance, ‘the President of the United States’ may denote Bush with respect to a given possible world and a certain time, but Bill Clinton with respect to that world and another time span.

Let me refer to the semantic features for the aforementioned expressions as their semantic value: so, the semantic value of a sentence (with respect to a possible world and/or a time) is its truth value (with respect to that world and that time); the semantic value of a definite description (with respect to a possible world and/or a time) is its *denotatum* (with respect to that world and that time); and so on. In general, and leaving momentarily aside a few interesting complications, an expression is rigid with respect to a parameter *k* if it has the same semantic value across all *ks*. For instance, an eternally true sentence is temporally rigid: its truth value is constant across all times. Since in what follows I leave temporal considerations out of the picture, I employ the expression ‘rigid’ *tout court* as alluding to constancy of semantic value across possible worlds.

Names and Rigidity

Two kinds of singular designators, namely names and definite descriptions, have been singled out as paradigmatic of the rigid/nonrigid divide. As I hinted, typical definite descriptions such as ‘the President of the United States’ are nonrigidly associated with distinct *denotata* with respect to different possible worlds. But proper names such as ‘George Bush’ apparently rigidly persist in referring to Bush with respect to all possible worlds (or, at least, all possible worlds in which Bush exists; more on this *caveat* later). For instance, if you are interested in evaluating ‘George Bush is wise’ with respect to a counterfactual scenario as dramatically different from the actual world as you may like, what you need to consider is whether, given the collection of individuals that are wise in that scenario, Bush is one of them (see Kripke, 1980).

One possible source of confusion must be cleared from the outset: what is at issue is not the question of who happens to be called ‘George Bush’ in the scenarios under consideration: that in some merely possible world certain expressions may be endowed with a different meaning, and certain names may be used to refer to alternative individuals is of no relevance for our purpose. Similarly irrelevant is the possibility that, in our actual employment of ‘George Bush’, different individuals may be referred to by that name type.

Types of Rigidity

I wrote that typical definite descriptions behave non-rigidly. This does not entail that there may not be rigid definite descriptions. For instance, ‘the positive square root of four’ denotes the number two with respect to any possible world and is thus a rigid designator. It seems clear that this expression’s rigidity derives not directly from its semantic behavior, but from the modal perseverance of the mathematical world. The description ‘the positive square root of four’ interacts with the parameter with respect to which it is evaluated as any other description would: a certain condition is being put forth, in this case that of being the positive square root of four, and an individual is selected on the basis of its ability to satisfy that condition. The description is, in a sense, semantically available for a nonrigid profile. Its rigidity is a result not of its semantic behavior, but of the response that alternative possible worlds provide: one and the same item, the number two, is inevitably being offered as the description’s *denotatum*.

Saul Kripke refers to descriptions such as ‘the positive square root of four’ as rigid only *de facto*. Expressions such as proper names, namely expressions whose semantic profile presumably guarantees their rigidity, are classified as rigid *de jure* (Kripke, 1980). Note that in this understanding of the *de facto/de jure* distinction, not all rigid descriptions are rigid merely *de facto*. For instance, the description ‘the actual President of the United States’ rigidly refers to a certain individual by virtue of its semantic behavior, in particular by virtue of the meaning of the indexical expression ‘actual’ (see Kaplan, 1977; Lewis, 1970).

Thus far, I rested satisfied with the notion that a rigid expression has a constant semantic value across possible worlds, and, in particular, that a rigid designator designates the same individual with respect to any possible world in which that object exists (Kripke, 1980; see also Kaplan, 1977). When it comes to the decision about an expression’s semantic value with respect to possible worlds in which that object fails to exist, two options may be entertained. In the terminology from Salmon (1981), a designator is **persistently** rigid iff it designates its *designatum* *x* with respect to any world in which *x* exists and designates nothing with respect to any world in which *x* does not exist. On the other hand, a designator is **obstinately** rigid iff it designates *x* with respect to every possible world, regardless of whether *x* does or does not exist in it.

Rigidity has played an important role in one of the central semantic debates during the seventies and

eighties, having to do with descriptivist views of customary proper names, such as, presumably, Frege’s and Russell’s. If proper names are rigid but typical descriptions are not, so the story goes, proper names may not behave in the manner descriptivists envisage. (For a critical discussion of a descriptivist reply, to the effect that proper names are associated with rigid descriptions, see Salmon, 1981; Soames, 1998.) Other important themes linked to the notion of rigidity have to do with the hypothesis that natural kind terms, such as *tiger* or *gold*, also behave rigidly (see Kripke, 1980; for a detailed discussion of the difficulties in extending the classic definition of rigidity to kind terms, see Soames, 2002). Finally, the idea of rigidity and the distinct but related notion of direct reference occupy center stage in the semantic analysis of indexical expressions. (On the relationships between rigidity and direct reference, see Marti, 2003.) For instance, an indexical singular term such as *I* rigidly refers (with respect to a given context *c*) to the agent of *c*, that is, at least in typical instances, to the person who is speaking. The notion that indexicals such as *I* behave rigidly, even though they are associated with a descriptive meaning, plays a crucial role in the classic arguments for the distinction between meaning and content or, in David Kaplan’s classic vocabulary, between *character* and *content* (see Kaplan, 1977; Lewis, 1980).

See also: Descriptions, Definite and Indefinite: Philosophical Aspects; Modal Logic; Possible Worlds: Philosophical Theories.

Bibliography

- Kaplan D (1977). ‘Demonstratives.’ In Almog J, Perry J & Wettstein H (eds.) *Themes from Kaplan*. New York and Oxford: Oxford University Press.
- Lewis D (1970). ‘Anselm and actuality.’ *Nous* 4, 175–188.
- Kripke S (1980). *Naming and necessity* (2nd edn.). Cambridge, MA: Harvard University Press.
- Lewis D (1980). ‘Index, context, and content.’ In Kanger S & Öhman S (eds.) *Philosophy and grammar*. Dordrecht: Reidel.
- Marti G (2003). ‘A question of rigidity in new theories of reference.’ *Nous* 37, 161–179.
- Salmon N (1981). *Reference and essence*. Princeton, NJ: Princeton University Press.
- Soames S (1998). ‘The modal argument: wide scope and rigidified descriptions.’ *Nous* 32, 1–22.
- Soames S (2002). *Beyond rigidity. The unfinished semantic agenda of naming and necessity*. Oxford: Oxford University Press.

Rules and Rule-Following

A Miller, Macquarie University, Sydney, NSW, Australia

© 2006 Elsevier Ltd. All rights reserved.

Why are the notions of rules and rule-following of significance to the philosophy of language and linguistics? Inspired by the discussions of rule-following in Wittgenstein's *Philosophical investigations* (sections 138–242) and *Remarks on the foundations of mathematics* (section VI), Saul Kripke and Crispin Wright independently developed arguments that challenged our intuitive conception of *meaning* as both *factual* and *objective* (see Wright [1980], Chapters 2 and 12, and Kripke [1982]). In this article, we will focus on the rule-following issue as presented by Kripke.

Rule-Following and Meaning

Suppose that Jones intends to follow the rule 'Add 2' in continuing the series: 2, 4, 6, 8, 10, Jones can continue the series in ways that *accord* or *fail to accord* with the requirements of the rule. For example, the continuation 12, 14, 16, . . . would accord with the rule (as standardly understood), whereas the continuation 12, 13, 14, . . . would fail to accord with the rule (as standardly understood). In other words, the former continuation would be *correct* by the lights of the rule, whereas the latter would be *incorrect* by the lights of the rule: the rule provides a *normative standard* against which particular continuations can be assessed as correct or incorrect. Intending to follow a rule is analogous in certain respects to meaning something by a linguistic expression. The meaning of a linguistic expression provides a normative standard against which the uses of that expression can be assessed as correct (or incorrect), as according with that meaning (or failing to accord with that meaning). Thus, for example, given the usual meanings of the numerals and signs involved, and given the arithmetical fact that the sum of 68 and 57 is 125, the answer '125' to the query '68 + 57 = ?' will be correct by the lights of those meanings, whereas the answer '5' will be incorrect by the lights of those meanings.

Suppose that Jones has laboriously written out the series 2, 4, 6, 8, . . . , 996, 998. Is there a fact of the matter as to what rule he's been following in writing out this sequence? Ordinarily, we'd be inclined to say that there is a fact of the matter, and, that, in all likelihood, he's been following the rule 'Add 2' (and not, say, the rule 'add 3,' or not following any rule at all but simply writing down numerals at random). According to Kripke's Wittgenstein

(hereafter 'KW'), this is an illusion: there is no fact of the matter as to what rule Jones has been following, and indeed no fact of the matter as to whether Jones has been following any rule at all. In similar fashion, we can ask whether there is a fact of the matter as to what meaning Jones associates with the '+' sign. Assuming that Jones's uses of the '+' sign have been roughly similar to our own, we'd be inclined ordinarily to say that there is a fact of the matter, and, that in all likelihood, he understands '+' to mean the *addition* function (and not, say, the multiplication function or no function whatsoever). As in the case of rules, KW claims that this is an illusion: there is no fact of the matter as to what Jones means by '+,' and indeed no fact of the matter as to whether he means anything at all by '+.'

Constitutive and Epistemological Skepticism

Before turning to KW's argument, we need to note a distinction between two types of skepticism. An *epistemological skeptic* about a particular area claims that we are not entitled to the knowledge claims that we typically make within it. Thus, for example, an epistemological skeptic about the external world (such as the skeptic who is Descartes's protagonist in his *Meditations*) would argue that I am not entitled to claim that I know that I am currently awake and sitting at a computer writing a philosophy essay. The epistemological skeptic doesn't question whether there is a fact of the matter as to whether I'm currently awake: he's happy to concede that there is a fact of the matter (either I'm awake, etc., or I'm not), and questions only whether I'm entitled to say that I know which of the two relevant possibilities is actually the case. KW's *skepticism about rule-following and meaning is not only epistemological skepticism*. KW argues, not only that we don't know whether we mean addition by '+,' or intend to follow the rule 'Add 2,' but also that there may not be anything to know. That is, not merely are we ignorant about what '+' means; there is no fact of the matter as to what it means – there is nothing for meaning facts to consist in or to be constituted out of. KW is thus a *constitutive skeptic* about rule-following and meaning. But as we'll see, KW takes an epistemological route to his constitutive skepticism.

KW's Skeptical Argument

Suppose that, in the examples above, Jones has never previously gone beyond 996 in his apparent attempts

to continue the arithmetical series generated by the rule 'Add 2,' and has never performed arithmetical calculations with numbers greater than 57 (the fact that Jones is a finite creature ensures that we can always set up this sort of scenario simply by choosing numbers that are large enough). If there were a fact of the matter as to what rule Jones is trying to follow, or a fact of the matter as to what meaning he associates with the '+' sign, there would be a fact of the matter as to how he ought to continue the number series (if he is to accord with the relevant rule), and a fact of the matter as to how he ought to answer the query ' $68 + 57 = ?$ ' (if his practice is to accord with the relevant meaning). KW argues that there is no fact of the matter as to how he ought to continue the number series, and no fact of the matter as to how he ought to answer the query ' $68 + 57 = ?$ '. Who is to say that Jones does not, in the case of the number series, intend to follow the rule 'Zadd 2,' in which the result of zadding 2 to a number is the same as the result of adding 2 to a number for numbers less than 1,000 and the same as the result of adding 4 to a number for numbers greater than or equal to 1,000? If Jones intended to follow this non-standard rule in continuing the series, he ought to continue the series 998, 1000, 1004, 1008, ... and not as we'd expect, namely 998, 1000, 1002, 1004, 1006 ... and so on. Likewise, who is to say that Jones does not, in the case of the arithmetical calculation, understand '+' to mean, not the addition function but, rather, the *quaddition* function, in which the result of *quadding* two numbers x and y is the same as adding them when both x and y are less than 57 but, rather, 5, when either x or y is greater than or equal to 57? If Jones meant the quaddition function by the '+' sign, he ought to answer '5' to the query ' $68 + 57 = ?$,' and not, as we'd expect, '125.'

KW's argument proceeds by allowing us *unlimited and omniscient access* to two broad categories of fact, and invites us to find the fact that constitutes Jones's intending to follow the 'Add 2' rule as opposed to the 'Zadd 2' rule, or Jones's meaning addition as opposed to quaddition by '+': facts about Jones's previous behavior and behavioral dispositions and facts about Jones's mental history and mental states. The assumption is that if facts about rule-following and meaning are to be found anywhere, they will be found within these two kinds of fact. So, if unlimited and omniscient access to these two classes of fact fails to turn up a fact that constitutes Jones's intending to follow one rule rather than another, or meaning one thing rather than another, it will follow that there is simply no such fact to be found. This is why we said above that KW's argument for constitutive skepticism proceeds via an epistemological route: KW argues

that even given unlimited and omniscient access to facts about Jones's behavior and behavioral dispositions and facts about Jones's mental life, we still couldn't claim to know, or justify, a particular hypothesis about how Jones ought to continue the arithmetical series or answer the arithmetical query.

Facts about Jones's previous behavior won't do the trick, as all of that behavior is consistent with Jones's intending to follow the 'Zadd 2' rule or meaning the quaddition function by the '+' sign: *ex hypothesi*, he had never previously dealt with cases in which the numbers were large enough for 'Add 2' and 'Zadd 2,' or meaning addition and quaddition, to demand different responses (Kripke, 1982: 7–15). Facts about how Jones is disposed to continue the series, or answer arithmetical queries, won't turn the trick either. Even if it was true that *if* Jones had reached 1000, he *would* have continued 1004, 1008, and so on, or true that *if* he'd been asked ' $68 + 57 = ?$ ' he *would* have answered '125' and not '5,' neither of these facts could plausibly be said to constitute his intending to follow one rule rather than another, or mean one function rather than another. Facts about the rule he intended to follow, or the meaning he attached, are *normative* facts, facts about how he *ought* to continue the number series or answer the arithmetical query. But the dispositional facts canvassed tell us at most what he *would* do in certain situations, as opposed to telling us what he *ought* to do in those situations (Kripke, 1982: 22–38). KW also rules out the following as facts constitutive of rule-following and meaning: *general thoughts* or *instructions* (1982: 15–17); relative *simplicity* of hypotheses about rule-following and meaning (1982: 38–40); *qualitative, introspectible, irreducibly mental states* (such as mental images) (1982: 41–51); *sui generis and irreducible mental states* (1982: 51–53); relations to *objective, Fregean senses* (1982: 53–54). Thus, according to KW there are no facts about rule-following and meaning, and these notions vanish 'into thin air' (1982: 22).

KW's Skeptical Solution

KW describes this result as a 'sceptical paradox,' and attempts to avoid the 'insane and intolerable' (1982: 60) conclusion that 'all language is meaningless' (1982: 71) via a 'sceptical solution.' The main idea of the skeptical solution is that judgments about rule-following and ascriptions of meaning can be viewed as possessing a *non-fact-stating role*, so that the practices of ours involving the notions of rule-following and meaning avoid the worries associated with the skeptical paradox. If the function of judgments about meaning is not to describe facts, the discovery that

there are no facts about meaning no longer threatens our practice of making such judgments. It emerges as a consequence of the detail of the skeptical solution that the notions of rule-following and meaning only have application relative to *communities* of rule-followers or speakers, not absolutely. In particular, it follows that 'solitary' language is impossible. (This is KW's take on Wittgenstein's famous 'private language argument'.)

Significance of the Issue

Unless it is blocked, the skeptical paradox threatens to undermine the idea that we can give a *cognitive-psychological* explanation of, for example, semantic creativity: the capacity speakers' have to understand novel utterances. One way of essaying such an explanation would be in terms of speakers' knowledge of the meanings of the familiar constituents of the novel utterance and their proceeding from this to knowledge of the utterance's meaning in a way that mirrors the compositional route to the meaning of the utterance from the meanings of its constituents and their mode of syntactic combination. Clearly, if there are no facts about meaning there will be no facts about the meanings of novel utterances or their constituents, so that the whole project of cognitive-psychological explanation of semantic creativity will be undermined. That would be a consequence of the skeptical paradox about meaning. But the skeptical paradox about rule-following in general would have even more destructive consequences: one way of attempting to explain the capacity speakers have to distinguish grammatical from ungrammatical strings would proceed in terms of their implicit grasp of grammatical rules governing the language in question. Clearly, if there are no facts of the matter as to whether these rules classify a given novel string as grammatical or not, this whole project, too, will be thrown in jeopardy. For a particularly clear exploration of these issues, see Wright (1989). See also Baker and Hacker (1984).

Responses to the Skeptical Paradox and Skeptical Solution

For defenses of dispositionalist, or otherwise naturalistic, solutions to KW's skeptical paradox, see Forbes (1984), Fodor (1990), Millikan (1990), and Horwich (1998). For further discussion of dispositionalism, see Boghossian (1989: Section V), McManus (2000), and Miller (2003). For non-reductionist responses (according to which facts about meaning are not reducible to facts about non-semantically and

non-intentionally characterized behavior or states), see McDowell (1984), McGinn (1984), Wright (1989), Boghossian (1989: Section VI), and Pettit (1990). Wright takes the rule-following arguments to threaten, not the factuality of meaning but, rather, the *objectivity* of meaning: see the Introduction to Wright (1993). For discussion of the skeptical solution and the argument against solitary language, see Blackburn (1984), Wright (1984), and Boghossian (1989: Sections II–IV). KW's skeptical solution was characterized above as a version of non-factualism about ascriptions of meaning, but this interpretation has been challenged by a number of philosophers: see, e.g., Wilson (1994) and Kusch (in press). For a searching examination of the normativity of meaning, see Hattiangadi (in press). For a reaction to KW from the perspective of a linguist, see Chomsky (1986). For an introductory survey and overview, see Miller (1998: Chaps. 5 and 6).

See also: Indeterminacy, Semantic; Normativity; Private Language Argument; Radical Interpretation, Translation and Interpretationalism; Use Theories of Meaning.

Bibliography

- Baker G & Hacker P (1984). *Language, sense and nonsense*. Oxford: Blackwell.
- Blackburn S (1984). 'The individual strikes back.' *Synthese* 58, 281–301. Reprinted in Miller and Wright (2002).
- Boghossian P (1989). 'The rule-following considerations.' *Mind* 98, 507–549. Reprinted in Miller and Wright (2002).
- Chomsky N (1986). *Knowledge of language*. New York: Praeger.
- Fodor J (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.
- Forbes G (1984). 'Scepticism and semantic knowledge.' *Proceedings of the Aristotelian Society* Supplementary Volume, 223–237. Reprinted in Miller and Wright (2002).
- Hattiangadi A (in press). 'The normativity of meaning.' *Mind and Language*.
- Horwich P (1998). *Meaning*. Oxford: Oxford University Press.
- Kripke S (1982). *Wittgenstein on rules and private language*. Oxford: Blackwell.
- Kusch, M. (in press). *A defence of Kripke's Wittgenstein*. Chesham: Acumen.
- McDowell J (1984). 'Wittgenstein on following a rule.' *Synthese* 58, 325–363. Reprinted in Miller and Wright (2002).
- McGinn C (1984). *Wittgenstein on meaning*. Oxford: Blackwell.
- McManus D (2000). 'Boghossian, Miller, and Lewis on dispositional theories of meaning.' *Mind and Language* 15, 393–399.

- Miller A (1998). *Philosophy of language*. London: UCL/Routledge.
- Miller A (2003). 'Does belief-holism show that reductive dispositionalism about content could not be true?' *Proceedings of the Aristotelian Society* Supplementary Volume, 73–90.
- Miller A & Wright C (eds.) (2002). *Rule-following and meaning*. Chesham: Acumen.
- Millikan R (1990). 'Truth rules, hoverflies, and the Kripke-Wittgenstein paradox.' *The Philosophical Review* 99, 323–353. Reprinted in Miller and Wright (2002).
- Pettit P (1990). 'The reality of rule-following.' *Mind* 99, 1–21. Reprinted in Miller and Wright.
- Wilson G (1994). 'Kripke on Wittgenstein on normativity.' *Midwest Studies in Philosophy* 19, 366–390. Reprinted in Miller and Wright (2002).
- Wittgenstein L (1953). *Philosophical investigations*. Oxford: Blackwell.
- Wittgenstein L (1974). *Remarks on the foundations of mathematics*. Oxford: Blackwell.
- Wright C (1980). *Wittgenstein on the foundations of mathematics*. London: Duckworth.
- Wright C (1984). 'Kripke's account of the argument against private language.' *Journal of Philosophy* 81, 759–778. Reprinted in Wright (2001).
- Wright C (1989). 'Wittgenstein's rule-following considerations and the central project of theoretical linguistics.' In George A (ed.) *Reflections on Chomsky*. Oxford: Blackwell. 203–239. Reprinted in Wright (2001).
- Wright C (1993). *Realism, meaning and truth* (2nd edn.). Oxford: Blackwell.
- Wright C (2001). *Rails to infinity*. Cambridge, MA: Harvard University Press.
-

S

Saussure: Theory of the Sign

P Cobley, London Metropolitan University, London, UK

© 2006 Elsevier Ltd. All rights reserved.

The sign theory of Ferdinand de Saussure (1857–1913) is principally to be found in his *Cours de linguistique générale* (*Course in General Linguistics* 1916; translated into English in 1959 and 1983). Saussure's previous work in philology and his focus on the development of Indo-European languages necessarily contains ruminations on the vicissitudes of the linguistic sign but it was his delivery of the course in general linguistics, at short notice to replace a sick colleague, at the University of Geneva from 1907 to 1911, which established his thought in this area. His posthumously published treatise of 1916 was based not on his own writings but on the notes of students present at the course. Specifically, the course was 'reconstructed' for publication by Charles Bally and Albert Sechehaye who stated in their introduction that they were aware of their responsibility to the author who "perhaps might not have authorized the publication of this text" (Saussure, 1983: xx).

The preliminary fact that needs to be taken into account, then, in relation to Saussurean sign theory is that, from the outset, it was never definitive. Readers of the *Cours* cannot be confident that Saussure's exact words are reproduced in full. Furthermore, they cannot be sure that the editorial choices of Bally and Sechehaye were not limited by what was available to them. For example, a notebook of notes in Saussure's hand, unknown to the editors, was discovered in 1996 and may force a re-reading of previous understandings of Saussurean sign theory. More important still, perhaps, is that Saussure's work has been refracted through the work of many illustrious successors beyond the realms of academic linguistics. While Saussure's *Cours* is just one, somewhat fragmented, publication that is still insufficiently read by those dabbling in sign study, the publications and teachings of his successors (or those who have invoked his name) have reached far wider audiences. The latter have had the benefit of an age of media saturation as well as a growing preoccupation with

the aspects of communication that lie beyond the objects of linguistics. It is difficult to talk about Saussure's sign theory, then, without mentioning that there is also the Saussurean sign theory of Karcevskij, Vološinov, Jakobson, Benveniste, Hjelm-slev, Lévi-Strauss, Lacan, Barthes, Mounin, Chomsky, Derrida, and Baudrillard, to name but a few. (Incisive discussions of a number of these are to be found in Harris, 2003).

Saussure's *Cours* projects "a science *which studies the role of signs as part of social life*" (Saussure, 1983: 15) early in the introduction, but the focus on the linguistic sign appears in Part One on General Principles. From the beginning, Saussure makes it clear that a linguistic sign is not "a link between a thing and a name, but between a concept and a sound pattern" (Saussure, 1983: 66). Moreover, "the sound pattern is not actually a sound; for a sound is something physical" (Saussure, 1983: 66). He therefore stresses that "This sound pattern may be called a 'material' element only in that it is the representation of our sensory impressions" (Saussure, 1983: 66). The linguistic sign is thus a "two-sided psychological entity" (see Figure 1).

It should be clear, then, that Saussure is not pursuing the relation between a thing in the world and the way that it is designated, but a psychological entity that amounts to signhood.

Proceeding to refer to the sound pattern as the *signifiant* and the concept as the *signifié*, Saussure makes clear that the former is in the mind and represents sensory impressions of sound outside the mind, while the latter consists of an abstract formulation of phenomena in the world such as 'house,' 'white,' 'see' and so forth (Saussure 1983: 65, 101ff). Saussure insisted that there was a *signifié* bound to each *signifiant* but that the reasons for their binding was not natural or pre-ordained. This fact is the first of the two fundamental characteristics of the sign: that the relation in the sign is arbitrary. He writes,

There is no internal connexion, for example, between the idea 'sister' and the French sequence of sounds *s-ö-r* which acts as its *signifiant*. The same idea might as well

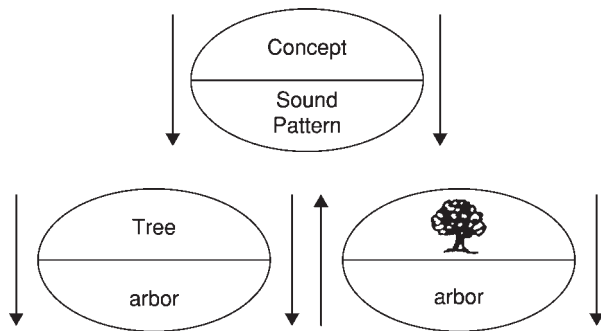


Figure 1 The linguistic sign as a “two-sided psychological entity” (cf. Saussure, 1983: 67).

be represented by any other sequence of sounds. This is demonstrated by differences between languages, and even by the existence of different languages. The signification ‘ox’ has as its signal *b-ö-f* on one side of the border [between French- and German-speaking regions] but *o-k-s* (Ochs) on the other side. (Saussure, 1983: 67–68; the terms *signifiant* and *signifié* are left in the original with the permission of the translator.)

To use another example, the connection between a *signifiant* ‘duck’ and the concept of ‘duckness’ was seen to be specific to the development of particular national languages. Nevertheless, Saussure did consider possible deviations from this principle, such as onomatopoeic words or signs such as those used in mime. These, somehow, seem to have a bearing which, to an extent, does suggest a preordained relationship of *signifiant* and *signifié*. But they are only deviations from the main rule. Tellingly, then, Saussure insisted that “when semiology is established . . . the main object of study in semiology will none the less be the class of systems based upon the arbitrary nature of the sign” (Saussure, 1983: 68).

The second principle upon which Saussure insisted was the linear character of the *signifiant*. This has not been the focus of semiology in the same way as the first principle has, but is nevertheless important. Saussure says that the linguistic *signifiant*, “being auditory in nature, has a temporal aspect” (Saussure, 1983: 69). What the principle of temporality entails is that the auditory sign can only exist in a linear fashion. Unlike graphic signs, it is always part of a chain unfolding. Pictures can present elements simultaneously; sounds have to wait for the connections to mount up one after another. Saussurean theories of the sign that attempted to look at nonverbal semiosis tended to elide this observation of Saussure.

Saussure insisted on four crucial factors inherent in the function of the sign:

- The arbitrary nature of the linguistic sign

- The great number of signs necessary to constitute a language
- The complex character of the system
- Collective inertia resists all linguistic innovations (Saussure, 1983: 73–74).

All of these bear upon whether the character of signs can be changed individually, a question Saussure answers in the negative. Arbitrariness seems to suggest that the sign allows free choice; however, Saussure emphasizes that words are fixed by our ancestors and that the effort required to change them is considerable and necessarily collective. This issue of freedom and its restrictions also bears upon some crucial distinctions in Saussure’s *Cours* that frame his theory of the sign. For Saussure, there is an emphatic difference between *langue* and *langage*. Rendering both terms accurately in English is difficult, but *langage* refers to a linguistic faculty of humans that *requires a langue* among its community of speakers. *Langage*, then, comprises two things. The first is the aforementioned *langue* – the differences between all signs (with their arbitrary relations), a set so multifarious and interconnected that it amounts to a system. This collective phenomenon, potentially available to, and stemming from, all “takes the form of a totality of imprints in everybody’s brain, rather like a dictionary of which each individual has a copy” (Saussure, 1983: 19). The second is *parole* – the sum total of what people say, and it comprises (a) individual combinations of words, depending on the will of the speakers, and (b) acts of phonation, which are also voluntary and necessary for the execution of the speakers’ combination of words” (Saussure, 1983: 19).

Saussure emphasizes that “there is nothing collective about speech” (Saussure, 1983: 19); *langue*, however, is precisely a matter of the collectivity, the system of differences shared by all speakers. *Langue*, in fact, is the system upon which the signs in individual speech acts is predicated. Harris writes:

The priority of *langue* as far as (Saussurean) linguistics is concerned can be boiled down to the proposition that if any episode of human speech is to be the subject of serious scientific inquiry it must be related in the first instance to a system which must be presupposed as underlying it (Harris, 2001: 121).

This is not the only way in which ‘choice’ over signs is subordinate to the system for Saussure.

Saussure indicates that speakers can combine linguistic signs into an infinite number ‘syntagmatic’ sequences, a fact that points to a certain amount of freedom. Such sequences – syntagmas, as Saussure calls them (Saussure, 1983: 121) – are bound by relations of combination. These relations exist to

facilitate the combinations **within** words and also **between groups of words**. (Within words, certain suffixes or prefixes might be subject to combinations, e.g., the suffixes *-ly* or *-less* with *name*; between words, certain combinations might be almost unavoidable, e.g., *dint* is very seldom seen without *by* preceding it and *of* succeeding it). Yet both are also subject to associative relations (Saussure, 1983: 123–124), or associations in the mind. Thus *-less* is not just a syntagmatic relation with *name*: it is also in an associative relation with all other uses of *-less* such as *blameless*, *feckless*, *careless*. Or, an associative relation may be in operation at the level of connectedness of concepts: *teaching*, *instruction*, *apprenticeship*, etc. (Saussure, 1983: 124). This has important consequences for the role of the sign in *langue* and *parole*. Saussure introduces the sentence, phrase, or clause as an example of a syntagma, but seems to hold back from suggesting they are always the subject of individual innovations. For Saussure, it seems that some combinations are a product of both *langue* and *parole*. Post-structuralists tend to suggest that ‘discourse’ (meaning many things, but frequently referring to the phrase, sentence, or larger syntagma) is an arena of choice or determination beyond the realms of *langue*. This is a matter that is implicitly taken up by a number of ‘Saussurean’ theories of the sign. What is apparent, though, in respect of the *Cours*, is that the status of *langue* as the fundamental system of signs is drawn into question by the possibility that it may be implicated with choice in syntagma production.

The role of the ‘value’ of the sign is clearer. Although combination is integral to Saussure’s sign theory, he insists that individual signs do not have ‘meaning’ or intrinsic identity. Rather, *langue* is a system of values: “A language is a system in which all the elements fit together, and in which the value of any one element depends on the simultaneous coexistence of all the others” (Saussure, 1983: 113). This is schematized in Figure 2.

Clearly, values arise from the relationships between whole signs. Yet these relationships are based on two key points:

Values always involve:

- something dissimilar that can be exchanged for the item whose value is under consideration, and

- similar things that can be compared with the items whose value is under consideration.

These two features are necessary for the existence of any value. To determine the value of a five-franc coin, for instance, what must be known is (1) that the coin can be exchanged for a certain quantity of something different, e.g., bread, and (2) that its value can be compared with another value in the same system, e.g., that of a one-franc coin, or of a coin belonging to another system (e.g., a dollar). Similarly, a word can be substituted for something dissimilar: an idea. At the same time it can be compared to something of like nature: another word (Saussure, 1983: 113–114).

Saussure is therefore keen to emphasize the system of values rather than the intrinsic meaning of signs. Thus, famously, he notes the value of the French word *mouton*, which carries both the animal to which it refers (the sheep) and the meat that comes from it (mutton); whereas the English word *sheep* has a different value because it needs to be supplemented by an additional word, *mutton*, to refer to the animal’s meat.

Saussure’s conclusion on value is that “In *langue* there are only differences, and no positive terms” (Saussure, 1983: 118). Putting this more broadly, he argues that “In a sign, what matters more than any idea or sound associated with it is what other signs surround it” (Saussure, 1983: 118). Effectively, if one were looking for it, this is the crux of Saussure’s sign theory and what imputes to it a particular character. One could argue that it is certainly the feature that seems to have drawn other ‘Saussurean’ sign theories to the *Cours* like a moth to a flame. Indeed, Saussure’s sign theory cannot really be considered without taking into account the many extensions, bowdlerizations, revisions, and plain distortions that it has suffered. Saussure’s *Cours* was rapidly taken up in many different ways, a fact that was not without its problems. Roy Harris writes:

Within five years of its publication Saussure’s *Cours* had become widely read in linguistic circles (De Mauro, 1972: 366). Translations into various languages followed. An initially critical reception gradually yielded to acceptance. By 1957 (centenary year of the eponymous Saussure’s birth) it was possible for a professional academic linguist to write: ‘We are all Saussureans now’ (Spence, 1957). But whether they all were – and, if so, to what extent – are tricky questions (Harris, 2001: 118).

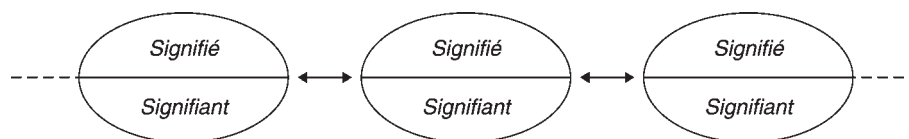


Figure 2 A system in which all the elements fit together (cf. Saussure, 1983: 113).

An early enthusiast for the *Cours* was the Russian, Sergej Karcevskij (1894–1955) who had attended Saussure's courses in Geneva and spread the word of their import back to Moscow in 1917. Saussurean sign theory was thus conveyed to the Moscow Linguistic Circle, featuring the young Roman Jakobson (1896–1982) and to OPOYAZ, a society that featured the Russian Formalists. One of the first major extensions of Saussure's theory of the sign, however, emanated from elsewhere in Russia. In fact, it was part of a penetrating critique by the Soviet theorist Valentin Vološinov. Vološinov (1895–1936), who had written a number of articles since 1922, published two books in the Soviet Union: in 1926, *Freudianism: a critical sketch* (1987) and, in 1929, *Marxism and the philosophy of language* (1973). Both remained virtually unknown for three decades except for the enthusiasm about the latter that was generated by Roman Jakobson in his communications to the Prague Linguistic Circle in the 1930s (Titunik, 1987: xv). At about the time that Vološinov was rediscovered in the Soviet Union, coinciding with the renaissance of Soviet semiotics (see Lucid, 1988) and leading to the translation into English of *Marxism and the philosophy of language* (1973), a number of sources began to argue that the author of Vološinov's works was none other than Mikhail Bakhtin (1895–1975) (see Todorov, 1984: 6–13).

Vološinov proceeds from an understanding of the linguistic sign as totally shot through with the social world from the outset. In fact, he devotes a considerable amount of time to criticizing "individualistic subjectivism" in the philosophy of language for simply studying words/utterances without the social dimension. His other target is the Geneva school (Saussure and his students, Bally and Sèchéhayé) as representative of "abstract objectivism." Saussure's *Cours* is 'abstract' because it does not take as its main focus concrete utterances, opting, instead, for 'form' (*langue*, the non-concrete system of differences) as the basis of language study. It is 'objectivist' because it promotes the idea that this form rules concrete utterances and is the common denominator of all meaning among all people. In general, the *Cours* is argued to err in suggesting that the social world is constituted in the shared system *langue*. For Vološinov, actual utterances (*parole*) must be understood in terms of the role played by a speaker and listener: "A word is a bridge thrown between myself and another . . . A word is a territory shared by both addresser and addressee, between the speaker and his [or her] interlocutor" (Vološinov, 1973:). All signs are built to be received, even if the identity of the receiver is unclear or he or she is not present; what

is important is that speaker and listener are bound in a social relationship that derives from both the utterance of signs and the wider set of social relationships in which the verbal interaction takes place. What is influential in Vološinov's *Marxism and the philosophy of language*, found also in Bakhtin, but arguably derived from dialogue with Saussure's *Cours*, is a 'dialogic' theory of the word.

Whereas Saussure sees instances of *parole* as being generated by a set of differences (in *langue*), Vološinov sees them as products of social situations. For Saussure, then, the production of utterances is ruled by purely formal laws while for Vološinov it is ruled by the multifarious nature of 'ideology.' 'Ideology' is to be understood here, as Vološinov makes clear at the beginning of *Marxism and the philosophy of language*, as both a part of reality and a reflection and refraction of a wider reality by means of signs (Vološinov, 1973: 9). Ideology is not to be understood as consciousness but as a sign system like that which makes up utterances; moreover, it fills the territory that is engendered by a verbal interaction. This, Vološinov argues, is because speakers and listeners are not concerned with the form of language, nor do they utter or hear 'words' as such; rather, they say and hear "what is true or false, good or bad, important or unimportant, pleasant or unpleasant, and so on" (Vološinov, 1973: 70). It is for this reason that any analysis of language must take into account the "common spatial purview of the interlocutors," "the interlocutors' common knowledge and understanding of the situation" and "their common evaluation of that situation" (see Vološinov, 1987).

In *Marxism and the philosophy of language*, Vološinov had also introduced 'theme,' 'meaning,' and accent,' concepts that indicate the material fashion in which meaning is different each time an utterance occurs because the situation is different on every occasion. The term 'meaning' is quite a problematic one for both Vološinov and Saussure: the latter insists on 'value' rather than 'meaning' as the distinguishing attribute possessed by a sign in a system of differences; likewise, Vološinov denies the possibility that a sign can have 'meaning' on its own. But whereas Saussure locates 'value' in the interaction of signs, Vološinov finds that 'meaning' is only part of a 'theme' that resides in the whole interaction between speakers. Moreover, 'meaning' is subject to the specific socially oriented intonation of utterances that Vološinov names 'accent.' When two people are in a room and the single word 'Well' passes between them, what does it mean? Saussure had maintained that the linguistic sign was an entity that was

collective by virtue of its existence in *langue*. Yet, the collectivity that enthuses the *Cours* was not enough for Vološinov, simply because it was ‘abstract’ rather than concrete.

Marxism and the philosophy of language contained a major critique of Saussurean sign theory but, nevertheless, can be argued to be an extension of it. In his positing of ‘theme’ and ‘meaning,’ the latter might be said to be concurring with the general principle of Saussure’s notion of *langue*. Similarly, Vološinov does not deny that there are instances of *parole* that are generated from existing potentials. However, there would always be one factor distinguishing Vološinov’s position from that of Saussure: for a sociolinguistic theory such as Vološinov’s, performative rules or a system of differences are insufficient for analyzing the fortunes of the linguistic sign (Vološinov, 1987: 99).

Although Vološinov’s critique of Saussure is often taken to be one of the first extensions of the latter’s sign theory that has impinged upon contemporary semiotics, the closely argued essays on linguistics of Emile Benveniste (1902–1976) were taking Saussure’s theory of the sign as their impetus. Benveniste was an Egyptian-born French linguist and defining figure in the thought of postwar France and beyond. Clearly, he is the father of post-structuralism, his work from the late 1930s onward paving the way for the critiques of structuralism offered by the likes of Derrida, Lacan, Kristeva, the later Barthes, Baudrillard, and assorted Anglo-American theorists in studies of film, literature, and philosophy (see Easthope, 1988). The essays in *Problèmes de linguistique générale* (1966, translated into English in 1971) ranged from a penetrating critique of Saussure’s principle of arbitrariness in the sign, ‘The nature of the linguistic sign,’ through a series of extensions of Saussurean sign theory. These include consideration of the general role of prepositions, ‘The sublogical system of prepositions in Latin,’ to his essay on the third person as a non-person, ‘The nature of pronouns.’

Benveniste argued in ‘The nature of the linguistic sign’ that the relations between the sound pattern of a sign and the mental concept attached to it were far from being arbitrary in experience. Where Saussure had insisted that the connection between a signifier and a signified was not a ‘natural’ one, rather it was based on convention or habit, Benveniste argued first that the arbitrary relation is one that exists between the whole sign and the object in the real world that it refers to. Relations between the *signifiant* and *signifié*, however, are not arbitrary at all for language users; in fact, the sound pattern and the concept almost seem to be as one in the human mind. In fact,

the creation of a concept in the human mind that is bound with a sound pattern is rehearsed so often at an early age that it becomes an almost instantaneous process. The individual sign user cannot change the relation in English between the *signifiant* of *duck* and its *signifié*. Consequently, Benveniste’s essay concludes, the connection between a sign’s sound pattern and the concept with which it is bound seems almost natural and should, more accurately, be stated as a necessary rather than an arbitrary relation.

‘The nature of the linguistic sign’ is characteristic of Benveniste’s take on linguistics: his extensions of Saussure are invariably conducted in an effort to apprehend the user’s feelings in language. Benveniste’s concern with the way in which signs work for the human subject points the way to further influential explorations of (post) Saussurean sign theory. In his essays on ‘Subjectivity in language’ and ‘The nature of pronouns’ collected in *Problèmes de linguistique générale*, Benveniste sought to demonstrate how signs under certain linguistic categories not only allow human subjects to refer to themselves but actually create the parameters of human self-consciousness. He writes,

‘Ego’ is he who says ‘ego’. That is where we see the foundation of ‘subjectivity’ which is determined by the linguistic status of ‘person’.

Consciousness of self is only possible if it is experienced by contrast. I use *I* only when I am speaking to someone who will be a *you* in my address (Benveniste, 1971: 225).

Because *I* is a linguistic sign – and therefore usable by the whole linguistic community – it cannot begin to represent the fullness of one human being’s self-consciousness. It is only a word. However, as relations in the linguistic sign are necessary, it invariably feels natural to humans that the word does have this power to represent.

In this light, it is easy to see how Benveniste so influenced post-structuralism. “It is in and through language that man constitutes himself as *subject*,” he writes, “because language alone establishes the concept of ‘ego’ in reality, in *its* reality which is that of being” (Benveniste, 1971: 224). For Benveniste, the separation of *I* and *you* in dialogue was crucial to the category of *person* because it is the means by which the individual sets himself or herself up as a subject in discourse. The personal pronouns are just one, albeit most important, means by which each speaker appropriates a language; deixis is another means, demanding that meaning can only be realized with reference to the instance of discourse in which the deictic category appears. As such, language creates the

designation of person; but it also contributes to the human understanding of such supposedly autonomous phenomena as time and space.

Benveniste recognizes in Saussure's theory of the sign a fundamental split that he extends to the realm of subjectivity. If the relations in the linguistic sign are 'arbitrary' when placed under analysis, but 'necessary' (felt to be natural and automatic) when used, it is possible that this indicates the subject of signs is pulled two ways. Benveniste identifies two sides of any use of language: he calls these *énoncé* and *énonciation*. The *énoncé* is simple enough: it is the statement or content of the particular instance of language, what is being said. The *énonciation*, on the other hand, is the act of utterance and presupposes a speaker and a listener. The two can be recognized when separated in this abstract way but, in practice, they are always entangled. The subject of this dynamic in linguistic signs is in a dilemma. There will be the rendering of himself or herself as a subject represented in the use of pronouns such as *I* (*énoncé*); but there will also be that other 'I' who does the rendering (*énonciation*). The dilemma, here, is made clear in such paradoxical constructions as *I am lying*, in which the subject speaking **must** be separate from the subject represented in the instance of discourse.

If Benveniste's furthering of Saussurean sign theory took place in relation to a theory of the subject, Louis Hjelmslev's contemporaneous co-opting of Saussure was more oriented to the extension of the system. A Danish linguist and member of the Copenhagen School of linguistics, Hjelmslev founded the post-Saussurean linguistics called glossematics. Effectively, glossematics extended Saussure's investigations into *la langue*, particularly the latter's insistence that this is a form and not a substance. Harris writes,

Glossematicians took the Saussurean dictum that *langue* is "form not substance" (Saussure, 1916: 163) to its logical conclusion, and argued that the languages now in existence and available to observation are merely historical realizations of certain systems that could equally well exist in other manifestations (not necessarily spoken and not necessarily written either). Thus "the task of the linguistic theoretician is not merely that of describing the actually present expression system, but of calculating what expression systems in general are possible as expression for a given content system, and vice versa" (Hjelmslev, 1970: 105) (Harris, 2001:128).

Whereas Hjelmslev had previously focused on distinctive features as elements of form in his early work, glossematics sought to identify systematic 'functioning' of signs. He embarked on an identification of signs in *langue* as, precisely, 'functions' in relation to 'functives.' A function exists, for example, between a

class and its components, a chain and its parts or a paradigm and its members (or between members or parts or components). A functive, on the other hand, is the 'terminal' – at either end – of a function.

Hjelmslev proceeds from the preliminary notions that language is a system of signs and that the sign – by virtue of its being for someone – must be construed as a function. However, he insists that the sign conceived in this way must be clarified as a 'sign-expression.' As such, then, there are two planes to be considered in the analysis of language as a sign system. The first is the 'expression plane,' which has both a form and a substance. The second is the 'content plane,' which has the same. Both are made up of a restricted number of non-sign entities, *figurae* (such as phonemes), which can combine to produce an unlimited number of signs. The fact that there are two planes of the sign means that, for Hjelmslev, the notion of the sign as simply for someone is limited. This view tends to see the sign as a matter of an indication by an 'expression' of some external 'content,' whereas the Saussurean view, according to Hjelmslev, sees the sign as generated by the connection between an expression and content. If the sign relation is seen as a function, however, the functives of expression and content can never be separable and this has consequences for the conception of the sign. Chiefly, Hjelmslev's work contends that the sign does not just indicate; rather it is made up of 'content substance' ordered into 'content form,' is juxtaposed for sign users with other examples of content form (e.g., *ring* might indicate both jewellery for the finger and something that emanates from the telephone), and is constituted also on the plane of expression as an 'expression form' (containing specifically organized *figurae*), which is itself ordered to an 'expression substance' (other pronunciations, other enunciations by other people, etc.).

Hjelmslev's glossematics thus instituted a procedure that treated all text elements as relations of function within expression and content. Yet, because the relation of function was seen by him as all important, and because both were made of ('meaningless') *figurae*, those elements that make up 'expression' and those that make up 'content' are to be considered interchangeable. It was the combination of Hjelmslev's resolute systematization and his facilitation of the mutability of planes according to function that made his glossematics so amenable to other structuralist linguists and thinkers.

The early work of Roland Barthes (1915–1980) marks him, perhaps, as heir to Hjelmslev as much as to Saussure. This is clear in some of Barthes formulations such as 'denotation,' 'connotation,'

and ‘metalanguage.’ However, ultimately, Barthes saw himself as furthering Saussure’s sign theory and, certainly, initiating Saussure’s vaunted project of a semiology, a science that studies the life of signs within society. Commentators tend to agree that Barthes’s work developed significantly enough during his career for there to be a manifest difference between the early and later Barthes. Usually, it is assumed that the earlier Barthes is a semiologist concerned with Saussure’s sign theory, while the later breaks with this. However, it is not too difficult to see extensions and continuities of Saussure from *Mythologies* (1957) to *S/Z* (1970). Barthes book of 1964, *Elements of semiology* (translated into English, 1967a), is particularly noteworthy because it purports to be both a treatise and a primer of Saussurean theory of the sign. Indeed, *Elements of semiology* has been, in many instances, almost taken as the word of Saussure rather than Barthes.

For Barthes in *Elements of semiology*, it is the repetition of signs in successive discourses that entails that each sign become an element of the language (*la langue*, “language without speech” [Barthes, 1967a: 14–15]). However, according to Saussure, *langue* is a heuristic means for thinking about the language system; language without speech would actually be impossible. Barthes was keen to explicate an understanding of language not as a collection of discrete signs but as an organizing principle of discursive fields. Such discursive fields would not be constituted by verbal signs (and their connections) alone. Barthes therefore draws attention to the fashion system, where signs work in a somewhat different way. For Barthes, the garment cannot be an instance of speech; rather, it is always “a systematized set of signs and rules: it is a language in a pure state” (Barthes, 1967a: 26). This is because “fashion clothes (as written about) are the language at the level of vestimentary communication and speech at the level of verbal communication” (Barthes, 1967a: 26). The feature of the sign that seems to be most crucial for Barthes and that encourages him to make the counterpoint to Saussure is the idea of the sign as part of an ‘ideological’ organizing principle. Whereas individual or collective uses of speech (or individual or collective usages of food, the analogous example he uses; Barthes, 1967a: 28) might become part of language (or alimentary language) as a result of repetition or evolution, what makes up the constituents of ‘language’ in the case of fashion is the highly directed work of a ‘deciding group.’

Elements of semiology makes it apparent that central to his theory of the sign is the way that it can be not simply an ideological vehicle but, in fact,

ideological through and through. This is apparent in his comments on the nature of the *signifiant* and the *signifié* (usually translated into English, potentially confusingly, as ‘signifier’ and ‘signified’). Nevertheless, Barthes’s formulations in this area are the subject of some questions regarding their consistency and their Saussurean credentials. One reason for this is suggested by Roy Harris: that Barthes came to Saussure relatively late and that his understanding of the Saussurean sign was already refracted through the lens of his reading of Lévi-Strauss, Lacan, Jakobson, Martinet, Benveniste and, as noted above, Hjelmslev (Harris, 2003: 136). In respect of the latter, Barthes introduces accurate definitions of the *signifiant* and *signifié*, for example, but immediately reframes their operation in terms of form, substance, plane of expression, and plane of content (cf. Hjelmslev, 1970).

Barthes is initially very clear about the *signifiant* and *signifié*: they are an acoustic image and a mental concept, respectively. He underlines the latter, in fact, by writing “the signified [*signifié*] is not a thing but a mental representation of the thing” (Barthes, 1967a: 42). However, Harris points out that while Barthes’s formulations about the internal relations in the sign are unexceptional, his definition of what the sign (*signe*) is, concentrating on the relation between *signifiant* and *signifié*, misses its mark. For Harris, Barthes

fails to see that for Saussure the *signe* is indeed not only a linguistic unit but *the* linguistic unit, and not a mere ‘rapport’ between its constituent parts. *A fortiori*, the fundamental Saussurean tenet that in *la langue* semiological relations take priority over – and determine – units sinks without trace in Barthes’ exposition ... Barthes had not only missed the most original theoretical feature of Saussure’s account of linguistic structure, but reinstated precisely the concept that Saussure was at pains to reject, i.e., the sign as a mere correlation between antecedently given items (formal and semantic) (Harris, 2003: 141).

Yet Barthes’ placing of the sign before the semiological relations of *la langue* has a further twist that has been undoubtedly influential and has served to distort the picture for those who take the account in *Elements of semiology* to be the definitive exposition of Saussure. In drawing attention to the “nature of the signifier [*signifiant*],” Barthes writes that

it is purely a *relatum*, whose definition cannot be separated from that of the signified [*signifié*]. The only difference is that the signifier is a mediator: some matter is necessary to it. But on the one hand it is not sufficient to it, and on the other, in semiology, the signifier can, too, be relayed by a certain matter: that of the word. This materiality of the signifier makes it once more necessary

to distinguish clearly 'matter' from 'substance': a substance can be immaterial (in the case of the substance of the content); therefore, all one can say is that the substance of the signifier is always material (sounds, objects, images) (Barthes, 1967a: 47).

It is fairly clear why Barthes makes this un-Saussurean assertion about the *signifiant* and, indeed, he is not shy about it:

In semiology, where we shall have to deal with mixed systems in which different kinds of matter are involved (sound and image, object and writing, etc.), it may be appropriate to collect together all the signs, *inasmuch as they are borne by one and the same matter*, under the concept of the *typical sign*: the verbal sign, the graphic sign, the iconic sign, the gestural sign are all typical signs (1967a: 47).

For Barthes, the business of establishing a semiology in which verbal and nonverbal signs can be considered equally leads him to override Saussure's fundamental principle in which the relations within the linguistic sign are purely mental. This is probably the most characteristic feature of Barthes's theory of the sign, although, unfortunately, the fact that *Elements of semiology* has often been treated as a faithful explication of Saussure rather than as a manifesto, has prompted the erroneous belief that Saussure's *signifiant* is the same as Barthes' (see, for example, Bignell, 1997; Cobley, 1996).

Another feature of Barthes' revision of Saussure concerns the vaunted 'arbitrariness' of relations in the sign and, here, Barthes takes his cue from Benveniste and then Martinet. In respect of the word *ox*, Barthes suggests, after Benveniste, that the relation of *signifiant* and *signifié* "is by no means arbitrary (for no French person is free to modify it), indeed, it is, on the contrary, necessary" (Barthes, 1967a: 50). Ultimately, however, Barthes suggests that the relation is one that is determined by different degrees of 'motivation.' An unmotivated system, he argues, is one where signs are not founded by convention but by unilateral decision; so, for him, signs in *la langue* are not unmotivated but signs in the world of fashion (where there is an elite organizing body) are. Signs are motivated, on the other hand, when the relation between *signifiant* and *signifié* is analogical (Barthes, 1967a: 51). This is evidence of Barthes' attempt to smooth the transition from Saussure's linguistic sign to a general semiology containing, for example, pictorial signs whose *signifiant/signifié* relation is, frequently and unavoidably, one of motivation.

It is evidently important for Barthes that there be an analytic unity in his theory of both verbal and nonverbal signs, chiefly because the ravages of

ideology, when exposed, would be less convincing if they were distributed unevenly across different kinds of signifying system. Indeed, Barthes' justly famous essays on photography (1977a, 1977b) are important in this respect since they indicate that the denotative sign enacts a motivated relationship often as if it were in the service of 'validating' the injustice of the connotative sign, establishing its literalness and helping to ground ideology. Barthes's close analytic readings, many of them informed by an implicit use of Saussure's notion of syntagmatic-paradigmatic interaction, have informed the practice of students of media, culture, and communications to the present day, although the later work of Barthes is often assumed to transcend Saussurean sign theory as though his earlier work did not. In a later book, *S/Z* (Barthes, 1970), Barthes does consider signs from a slightly different angle. However, one could argue that this work still struggles with the vicissitudes of the Saussurean sign.

In the opening pages of *S/Z*, Barthes formulate two kinds of texts, the 'readerly' and the 'writerly,' where the reader is respectively an idle consumer of the signs in the text or a diligent producer of it, almost rewriting what is presented (Barthes, 1974: 3–4). Barthes goes so far as to actively advocate the latter position but, curiously, the intense analysis that follows these comments and makes up the majority of the book produces an implicitly conflicting conclusion. *S/Z* dissects a Balzac short story, 'Sarrasine,' dividing it up into very brief segments and by elaborating five codes through whose matrix the text passes. The result is to expose an abundance of meanings, making a text that might have been billed as a 'simple realist narrative' into one that now threatens to be eminently 'writerly.' It is not clear whether Barthes set out to demonstrate that 'realist' texts – exemplified by 'Sarrasine' – somehow carried signs that purported to possess a 'stable,' one-to-one *signifiant/signifié* relationship. Yet, if this was Barthes's intention, then it is clear that the final delivery of *S/Z* stymied it. The Saussurean version of the sign was by no means complicit with 'readerly' texts; indeed, the analysis might suggest that post-Structuralist versions of the sign were not guaranteed to produce 'writerly' texts.

If one feature of Barthes's analyses has been an extension of Saussurean sign theory to make it almost synonymous with the analysis of everyday phenomena (as opposed to an element in the understanding of linguistic systems), then his work is of a piece with that of the enormously influential structuralist anthropologist, Claude Lévi-Strauss (b. 1908). To be sure, Lévi-Strauss was heavily influenced by a

number of thinkers (Marx, Kant, Durkheim, Mauss, Jakobson) but most prominent among these, perhaps, was Saussure. Lévi-Strauss's work until the mid-20th century focused on kinship systems and marriage rules, while later he concentrated on belief systems embodied in myths and religion. Inspired by phonology, Lévi-Strauss had developed a methodology to investigate marriage alliances and myth. Yet, more specifically, crucial to Lévi-Strauss's conception of anthropology was Saussure's insistence on how the differentiability between linguistic signs gives rise to value (*valeur*).

Rather than carrying intrinsic 'meaning,' Saussure argued that linguistic terms were imbued with an identity that relied totally on their difference to other terms. For Lévi-Strauss, in turn, differentiability characterized a range of anthropological phenomena, particularly in the practices of exchange to be found in kinship rules, but also elsewhere. This included such binary formulations as the 'raw' and the 'cooked,' one of a number of anthropological oppositions discussed at length by Lévi-Strauss in a Saussurean spirit heavily refracted through the work of Jakobson and Troubetzkoy. Taking the oppositions instanced by phonemes as the smallest possible units participating in the differentiability of language, Lévi-Strauss proceeded to formulate both a study of myth and of 'savage' thought in general. The analogy of phonemes combining and opposing to produce linguistic signs that might be employed in systems of exchange informed, for example, his analysis of kinship.

In his theory of myth, Lévi-Strauss considered mythic narratives to rely on 'mythemes' – a term that directly echoes 'phonemes,' of course. The conclusions that Lévi-Strauss reached about Saussure – that "language can be analyzed into things which are at the same time similar yet different" (Lévi-Strauss, 1977: 209) – encouraged him to concentrate on the 'form' by which phonemes were combined rather than the strictly semantic orientations of a myth's content. Lévi-Strauss thus devised a method that enabled him to break down all myths to allow the identification of a basic substrate. The most famous example of this is his 1955 interpretation of the Oedipus myth in which he treats it "as an orchestra score would be if it were unwittingly considered as a unilinear series" (Lévi-Strauss, 1977: 213).

Say, for instance, we were confronted with a sequence of the type: 1, 2, 4, 7, 8, 1, 3, 4, 6, 8, 1, 4, 5, 7, 8, 1, 2, 5, 7, 3, 4, 5, 6, 8 ... the assignment being put to all the 1s together, all the 2s, the 3s, the result is a chart:

1	2		4			7	8
		2	3	4		6	8
1				4	5		7
1	2				5		7
			3	4	5	6	8

(Lévi-Strauss, 1977: 213)

He then goes on to break down the features of the Oedipus myth – for example, 'Oedipus marries his mother, Jocasta' and 'Antigone buries her brother, Polynices, despite prohibition' (Lévi-Strauss, 1977: 214) – which he reorders in a similar way. Effectively, Lévi-Strauss comes up with four columns of relations in the myth: (a) overrating of blood relations; (b) underrating of blood relations; (c) slaying of monsters; (d) (names for) difficulty of balance/standing. These mythemes suggest that the orchestrated myth has at its core a common human concern: anxiety over our origins. Put another way, the myth plays out the drama of humankind's quest to determine whether it has been forged in blood/earth or by human reproduction.

In the face of post-structuralism's revisions of his theoretical position and, in particular, the critique leveled by Jacques Derrida, Lévi-Strauss remained an unrepentant reductionist, revealing in his work the *langue* behind copious examples of *parole*. Lévi-Strauss was not alone in proposing such broad formulae at this time. In a series of articles in the late 1950s and early 1960s, which culminated in the 1966 book, *Sémantique structurale*, A. J. Greimas (1917–1992) worked out a semiotic project influenced by Hjelmslev and Saussure. Considering myth and narrative, Greimas, in a fashion that, from a distance at least, seems akin to that of Lévi-Strauss, gave priority to structural relations between narrative entities rather than to their intrinsic qualities. Lévi-Strauss, along with Vladimir Propp (1895–1970) and Greimas, effectively gave birth to the discipline known as 'narratology.' Narratology is a specific way of understanding narrative that was developed out of Russian Formalism but relied to a large extent on the theory of the sign put forth by Saussure. Sometimes 'narratology' is the name given to any form of analysis of narrative, but this is misleading. The term, instead, refers to a particular period in the history of narrative analysis that has had important consequences for other areas of study but that has also, itself, been transformed by other disciplines and perspectives. What characterizes narratology most readily is a systematic, thorough, and disinterested approach to the mechanics of narrative – an approach that provides a stark contrast to those that observe or seek out 'value' in some narratives (and not others) or that provide hierarchies of narratives

based on spurious categories such as the 'genius' of an author or artiste.

Roland Barthes's essay 'Introduction to the structural analysis of narratives' (Barthes, 1977, [1966]) and Tzvetan Todorov's *Grammaire du Décaméron* (1969), the latter of which actually coined the term '*narratologie*,' represented the birth of narratology proper. In the late 1960s and early 1970s, these paved the way for works by names frequently associated with the narratological enterprise: Mieke Bal, Seymour Chatman, Dorrit Cohn, Gérard Genette, and Gerald Prince. The fundamental distinctions in narratology between, for example, 'plot,' 'story,' and 'narrative,' were, in different ways, developed by the immediate precursors of the main narratological enterprise who came after Russian Formalism. Both Propp and Lévi-Strauss undertake analyses of narrative that focus on 'story' or, in the latter's case, in an extension of the linguistic terminology from Saussure, 'semantic structure.' In *Morphology of the folktale* (originally published in Russian in 1928, then auspiciously translated into English in 1958, appearing in book form 10 years later), Propp analyzed 100 Russian folk stories. Rather than attending to the surface differences of the narratives, he examined their underlying commonalities, particularly the basic function of actions in each story. Propp identified 31 functions characterizing the tales: for example, 'One of the members of a family absents himself from home'; 'An interdiction is addressed to the hero'; 'The interdiction is violated'; through to 'The villain is punished' and 'The hero marries and ascends the throne.' Each of these functions is, of course, carried out by one or more of the *dramatis personae*. Propp therefore proceeded to isolate the seven basic roles of characters in his sample, listing the particular sphere of action to which each belonged: the hero, the villain, the princess (sought-for) and her father, the dispatcher, the donor, the helper, and the false hero.

As differential signs in narrative, which Propp named 'functions,' it was easy to see the connection between such elements and Saussurean sign theory as adjusted by Hjelmslev. It is in this frame that the categories introduced by Greimas were so influential. Most importantly, Greimas emphasizes the functional nature of Propp's *dramatis personae* by referring, instead, to 'actants.' 'Actants' or 'actantial roles' are defined in the following ways: in relation to each other, in relation to their place in the narrative's 'spheres of action' or 'functions,' and in relation to their place in the logic of a narrative. In Greimas's revision of the *dramatis personae* in *Structural*

semantics (1983, [1966]), the actants comprise 'subject vs. object,' 'sender vs. receiver' and 'helper vs. opponent,' a set of categories that replaces Propp's analysis based on character roles (heroes vs. sought-for person, father/dispatcher vs. hero, helper/provider vs. villain/false hero). Narrative meaning in this formulation is played out through the various functions: thus, the 'subject' searches for the 'object'; the 'sender' is on a quest, initiated by a 'subject' for an 'object'; and so on. In addition, Greimas explores the way in which the structure of narrative can be defined by the interaction of 'positive' and 'negative' functions such as the opposition between a relationship designated *A* and containing 'command/behest' (*a*) and 'acceptance' (*non a*), and a relationship designated *A* • consisting of 'violation' (*non a* •) and 'interdiction' (*a* •). Ultimately, the systematic scrutiny of these relations led Greimas to posit a 'semiotic square' of such coordinates as a tool for the analysis of narrative meaning.

The work of Greimas, as well that of many others, such as Barthes, had traveled a long way since adopting broadly Saussurean principles. Indeed, the extensions of Saussurean sign theory were often actually extrapolations from Saussure as explicated in the work of other linguists or cultural theorists. Another case in point is the sign theory of Jacques Lacan (1901–1981), which is heavily reliant on the work of Jakobson, Lévi-Strauss, and, especially, Benveniste. For Lacan as a psychoanalyst, the chief point about the linguistic sign was that it somehow misses something; it is unable to express the purely personal. This is because relations in the sign are arbitrary, allowing the same sign to be used by everybody else. In 1957, Lacan delivered a lecture at the Sorbonne entitled 'L'instance de la lettre dans l'inconscient ou la raison depuis Freud,' which incorporates a significant extension of Saussure's sign theory, while also pursuing a series of related questions evident in Lacan's work since the completion of his Ph.D. dissertation in 1932 (see Benvenuto and Kennedy, 1986: 107; Payne, 1993: 74).

Taking the fundamental constituent of the *Course in general linguistics*, Saussure's map of the sign with *signifié* on top of *signifiant* enclosed by an ellipse, Lacan inverts it and opens it up. He presents a picture of two identical doors distinguished only by the writing that appears above them (Figure 3).

Clearly, this is a version of Saussure's map (or algorithm) of the sign (Figure 4).

What is important here is that the priorities of Saussure are reversed by Lacan. The *signifié* (or mental concept) of what lies behind each door is

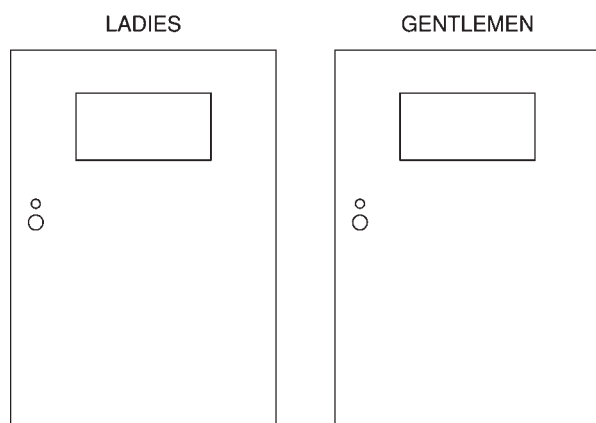


Figure 3 Lacan's twin doors. From Lacan, 1977.

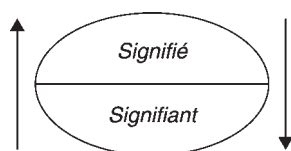


Figure 4 Saussure's sign. From Saussure, 1983: 113.

determined entirely by the *signifiant* (what should be the 'sound pattern,' here, is translated without comment by Lacan into a written notation: 'Ladies,' 'Gentlemen') that lies above the door. A whole cultural law – as Lacan puts it, the subjection of public life to the law of urinary segregation – is engendered by the existence and location of the two *signifiants* as well as their difference from one another. Lacan seeks to demonstrate that the mental concept is created by the way in which language operates. So, in addition to inverting the priorities in the algorithm, Lacan also banishes the ellipse and reinforces the bar that separates *signifiant* and *signifié* to emphasize that there is an incessant possibility of the *signifié* sliding under the *signifiant*. The *signifiant* is therefore responsible for the birth of subjectivity; however, it does not give birth to the human as a whole. There remains in the human a complex domain of desires – the unconscious – ruled by an order that is totally different from that of the social institution, language.

The implications of Lacan's inversion of the Saussurean algorithm for the study of subjectivity are fairly clear. As such, his work was taken up in all manner of areas in the humanities. It was especially influential in film theory circles, particularly in the 1970s in Britain. Problematically, however, it was frequently seen as correcting the fundamentals of Saussure's sign theory (or, as was frequently the

case, misrepresenting his theory of the sign in order to carry out some fixing that was never needed in the first place) or, worse, seen as standing for the truth of the Saussurean sign. This is a pattern that can be seen in other Saussurean sign theory, especially in the Gallic tradition – for example, Derrida (1976: 27–73) and Baudrillard (1983) – but also as Saussurean sign theory was used as a central figure in other disciplines (see, for example, Coward and Ellis, 1977; Culler, 1975; Hall *et al.*, 1980; Dyer, 1982; Fiske, 1990; Bignell, 1997; Hall, 1997; Cobley, 1996). Yet there is one area of Saussurean sign theory which has been of utmost importance to its dissemination and must be briefly mentioned here.

In his indispensable volume on *Saussure and his interpreters* (2003), Roy Harris discusses the Students' Saussure, the Editors' Saussure, Bloomfield's Saussure, Hjelmslev's Saussure, Jakobson's Saussure, Lévi-Strauss's Saussure, Barthes's Saussure, Chomsky's Saussure, Derrida's Saussure, and History's Saussure. Implicitly discussed, but it could also have merited a chapter of its own, would be the translators' Saussure. The English translation of the *Cours* is interesting in itself and, indeed, as Harris himself produced the second translation of the *Cours* in 1983, modesty may have prevented him from discussing the issue in his volume. However, it should be noted that the first English translation of 1959 has damaged the understanding of the *Cours* and caused it to be falsely applied in various areas. One example from Wade Baskin's translation of 1959 will suffice to illustrate the problem. It is the rendering of *signifiant*, *signifié*, and *signe* as 'signifier', 'signified,' and 'sign.' The first of these items produced the major misunderstanding. In short, it gave the impression to English natives that the *signifiant* was anything that did the work of signifying or, to put it another way, a sign. This was precisely the formulation that Saussure wanted to avoid and Saussure is explicit about this on a number of occasions in the *Cours*. The term for the *signifié*, at the same time, seemed to imply anything that was the object of signification – that is to say, any **thing** that was being **signified**, whether that be a thing in the head or a thing in the world (with the implication that it was generally the latter). At a stroke, Saussure's psychological conception of the sign was lost and versions of semiology were given free rein to look at the whole panoply of cultural artifacts as if they embodied a *signifié/signifiant* relationship. In his 1983 translation, Harris did his best to regain some of the ground that had been lost by rendering *signifiant* and *signifié* as 'signal' and 'signification' in an attempt to give a more accurate flavor

of Saussure's sign theory and to head off erroneous extrapolations. However, the damage was already done and Harris' terms have not been widely taken up as yet. (Thus the present article retains the original French terms).

What is beyond doubt is that Saussure's theory of the sign has been extraordinarily influential and has generated some productive work while also spawning some monsters. The posthumous provenance of the *Cours* has, perhaps, entailed that the 'real Saussurean sign theory' will never be known. However, two points should be added. The theory of the sign in Saussure's *Cours* is still unfolding as new discoveries about the original Geneva course in general linguistics are made (see Harris, 2003: 214–252). And, while it may not be the word of Saussure himself, Saussure's *Cours* is still insufficiently read.

Bibliography

- Barthes R (1964). *Éléments de sémiologie*. Paris: Seuil.
- Barthes R (1967a). *Elements of semiology*. Lavers A & Smith C (trans.) London: Cape.
- Barthes R (1967b). *Système de la mode*. Paris: Seuil.
- Barthes R (1970). *S/Z*. Paris: Seuil.
- Barthes R (1973). *Mythologies*. Lavers A (trans.). London: Paladin.
- Barthes R (1974). *S/Z*. Howard R (trans.). Oxford: Blackwell.
- Barthes R (1977, [1966]). 'Introduction to the structural analysis of narrative.' In Heath S (ed. & trans.). *Image-Music-Text*. Glasgow: Fontana.
- Barthes R (1977a). 'The rhetoric of the image.' In Heath S (ed. & trans.). *Image-Music-Text*. London: Fontana.
- Barthes R (1977b). 'The death of the author.' In Heath S (ed. & trans.). *Image-Music-Text*. London: Fontana.
- Baudrillard J (1983). *For a critique of the political economy of the sign*. Levin C (trans.). St. Louis: Telos.
- Benveniste E (1971). *Problems in general linguistics*. Meek M E (trans.). Coral Gables: University of Miami Press.
- Benvenuto B & Kennedy R (1986). *The works of Jacques Lacan*. London: Free Association.
- Bignell J (1997). *Media semiotics: an introduction*. Manchester: Manchester University Press.
- Cobley P (1996). 'Introduction.' In Cobley P (ed.) *The communication theory reader*. London: Routledge.
- Coward R & Ellis J (1977). *Language and materialism: developments in semiology and the theory of the subject*. London: RKP.
- Culler J (1975). *Structuralist poetics: structuralism, linguistics and the study of literature*. London: RKP.
- Derrida J (1976). *Of grammatology*. Spivak G C (trans.). Baltimore and London: Johns Hopkins University Press.
- Dyer G (1982). *Advertising as communication*. London: Routledge.
- Easthope A (1988). *British post-structuralism since 1968*. London: Routledge.
- Fiske J (1990). *Introduction to communications studies* (2nd edn.). London: Routledge.
- Greimas A J (1983, [1966]). *Structural semantics: an attempt at method*. McDowell D, Schleifer R & Velie A (trans.). Lincoln: University of Nebraska Press.
- Greimas A J (1987). *On meaning: selected writings on semiotic theory*. Perron P & Collins F H (trans.). Minneapolis: University of Minnesota Press.
- Greimas A J (1990). *The social sciences: a semiotic view*. Minneapolis: University of Minnesota Press.
- Hall S (1997). 'The work of representation.' In Hall (ed.) *Representation: cultural representations and signifying practices*. London and Thousand Oaks: Sage.
- Hall S et al. (eds.) (1980). *Culture, media, language*. London: Hutchinson.
- Harris R (2001). 'Linguistics after Saussure.' In Cobley P (ed.) *The Routledge companion to semiotics and linguistics*. London: Routledge.
- Harris R (2003). *Saussure and his interpreters* (2nd edn.). Edinburgh: Edinburgh University Press.
- Hjelmslev L (1970). *Language: an introduction ('Sproget')*. Whitfield F J (trans.). Madison and London: University of Wisconsin Press.
- Hjelmslev L (1970). *Prolegomena to a theory of language*. Whitfield F J (trans.). Madison and London: University of Wisconsin Press.
- Lévi-Strauss C (1977). *Structural anthropology 1*. Jacobson C & Grundfest Schoepf B (trans.). Harmondsworth: Penguin.
- Lévi-Strauss C (1987). *The view from afar*. Neugroschel J & Hoss P (trans.). Harmondsworth: Penguin.
- Lucid D P (ed.) (1988). *Soviet semiotics: an anthology*. Baltimore and London: Johns Hopkins University Press.
- Payne M (1993). *Reading theory: an introduction to Lacan, Derrida and Kristeva*. Oxford: Blackwell.
- Saussure F de (1983). *Course in general linguistics*. Harris R (trans.). London: Duckworth.
- Titunik I R (1987). 'Introduction.' In Vološinov V N (ed.) *Freudianism: a critical sketch*. Bloomington: Indiana University Press.
- Todorov T (1969). *Grammaire du Décaméron*. The Hague: Mouton.
- Vernant J-P (2001). *The universe, the gods and mortals: ancient Greek myths*. Asher L (trans.). London: Profile.
- Vološinov V N (1973). *Marxism and the philosophy of language*. New York: Seminar Press.
- Vološinov V N (1987). *Freudianism: a critical sketch*. Bloomington: Indiana University Press.

Scope and Binding: Semantic Aspects

A Szabolcsi, New York University, New York, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

The scope of an expression *E* is the domain within which *E*'s interpretation can impact that of other expressions. In the simplest cases this domain coincides with *E*'s sister constituent (*E*'s syntactic scope). The reading where *E*₁ scopes over some *E*₂ is notated as *E*₁ > *E*₂. *E*₁'s impact on *E*₂ may be of different kinds. Some examples follow. In (1) teachers may vary with boys; similarly in (2) if scope is used to effect pronoun binding. In (3) negation eliminates existential import. In (4) the Grail need not exist in the actual world. In (5) the polarity item is licensed.

- (1) Every boy saw a teacher.
'every boy > a teacher'
- (2) Every boy saw the teacher who flunked him.
'every boy > him'
- (3) Nobody drew a circle.
'nobody > a circle'
- (4) Parsifal seeks the Grail.
'seek > the Grail'
- (5) You have not responded yet.
'not > yet'

(6) summarizes Montague's (1974) classical technique for obtaining a reading where *everyone* takes wide scope, binding *his* and inducing referential variation in *a picture*. *Every boy* is interpreted as a generalized quantifier (6b), and a property is formed from the interpretation of its scope (6c). Pictures may vary because they are chosen relative to individuals playing the subject role. Pronoun binding occurs due to the identification of two arguments.

- (6a) *Every boy showed his father a picture* [Reading 1] is true if and only if the property (6c) is an element of the set (6b).
- (6b) *every boy*: the set of properties every boy has
- (6c) *_showed his father a picture*: the property of being an individual such that there exists some picture or other such that this individual showed that picture to this individual's father

(7) restates the right-hand side of (6a) using a logical notation similar to Montague's. Going beyond first-order logic, lambda is the abstraction operator and *P* a second-order variable over properties.

- (7) $\lambda P \forall x [\text{boy}'(x) \rightarrow P(x)] (\lambda y \exists z [\text{picture}'(z) \wedge \text{showed}'(y, \text{father-of}'(y), z)])$

Although the operator-variable notation is widely used, it is semantically not essential. For variable-free

semantics, see Szabolcsi (1989), Hendriks (1993), Jacobson (1999), and Kruijff and Oehrle (2003).

The intended scope of an expression is often not its sister in the most straightforward constituent structure of the sentence. For example:

- (8a) *Every boy showed his father a picture* [Reading 2] is true if and only if the property in (8c) is an element of the set (8b).
- (8b) *a picture*: the set of properties some picture or other has
- (8c) *every boy showed his father _*: the property of being an individual such that every boy showed this individual to that boy's father

Discrepancies are often resolvable by using a storage mechanism (Cooper, 1983), by positing a more abstract syntactic structure called Logical Form (May, 1985), or by making surface structure more articulated (Kayne, 1998). In other cases substantial semantic considerations or even new interpretive techniques must be invoked.

One semantically significant case is scope independence, exemplified by branching and cumulative readings (Schein, 1993; Hintikka and Sandu, 1997; Landman, 2000):

- (9) Some relative of every townsman and some relative of every villager hate each other.
- (10) Less than five detectives solved more than six crimes between them.

The main topic of this article is scope dependency. Many of the semantic innovations are necessitated by the observation that different operators have different scoping and binding abilities (Szabolcsi, 1997a). Nominal and adverbial quantifiers are the best studied. At least three classes emerge: (unmodified) indefinites like *some girl*, *two girls*, distributive universals like *every girl*, and counting quantifiers like *more/less than three girls*. Existential scope and distributive scope are distinguished, and E-type anaphora and covariance with situations are added to traditional bound variable and coreferential readings.

Thus, contrary to the traditional view, scope and binding are not uniform and are not unanalyzable primitives, and the two do not always go hand in hand.

Differential behavior regarding binding was observed first. (All referential dependencies involving pronouns will be indicated by underlining.) Indefinites, but not universals or counting quantifiers, support cross-sentential anaphora. Texts describing scenarios constitute an exception:

- (11) A girl is laughing. She is happy.
- (12) Every girl is laughing. # She is happy.

(13) More than one girl is laughing. # She is happy.

(14) Every graduate steps forward. The dean shakes hands with him or her.

Indefinites also support sentence-internal anaphora outside their syntactic scope. On the relevant reading both *a donkey* and *it* are within the scope of *every*.

(15) Every farmer who owns a donkey beats it.

These two binding peculiarities inspired varieties of dynamic semantics (Hintikka and Sandu, 1997; Heim, 1982; Kamp and Reyle, 1993; Groenendijk and Stokhof, 1991), in which indefinites are associated with referents that outlive the indefinite's syntactic scope. Alternatively, "donkey pronouns" are interpreted as definite descriptions (Evans, 1980; Neale, 1990), and donkey anaphora may be reinterpreted using situation semantics (Heim, 1990; Elbourne, 2001):

(15') 'every farmer who owns a donkey beats the donkey(s) he owns'

(15'') 'every minimal situation containing a farmer and a donkey he owns extends to one where the farmer beats the donkey'

Büring (2004) extends "situation binding" to other cases where a quantifier binds a pronoun it does not c-command:

(16) Every boy's mother likes him.

(17) Some person from every city likes it.

Indefinites are also special in that their scope within the sentence may escape the containing clause or even island:

(18) Every boy made up a story in which two girls had a bad dream.
'there are two girls such that every boy made up a story in which they had a bad dream'

But such extra-wide scope of indefinites only pertains to their ability to remain referentially independent of higher operators. Like universals and counters, plural indefinites cannot induce referential dependency in higher material. Boys cannot vary with girls in either (19) or (20). Only bad dreams can.

(19) Some boy or other made up a story in which every girl/less than three girls had a bad dream.

* 'for every girl/less than three girls there is a possibly different boy who ...'

(20) Some boy or other made up a story in which two girls had a bad dream.

* 'for each of two girls there is a possibly different boy who ...'

This duality cannot be expressed in traditional terms. It necessitates a distinction between existential scope, potentially unbounded, and distributive scope, typically limited to the smallest tensed clause (Beghelli *et al.*, 1997; Reinhart, 1997). Wide existential scope is attributed to referentiality (Fodor and Sag, 1982) or to the existential closure of a choice function variable (Egli and von Stechow, 1995; Reinhart, 1997). A choice function looks at sets and chooses one element or subset of each. According to Reinhart, the determiner in indefinites is interpreted as a choice function variable, whereas the numeral is a cardinality marker. *O_{choice} two girls* picks out a set of two girls. Distributivity, if present, is due to the interpretation of the clausemate predicate, e.g., [*each*] *had a bad dream*.

Extra-wide scoping indefinites may nevertheless be referentially dependent:

(21) Every professor praised every student who reviewed a certain book (that he wrote).
'every professor > a certain book > every student'

Such dependencies have been attributed to intermediate existential closure or to an overt or understood bound pronoun in a maximal-scope indefinite (Skolemized choice functions) (Kratzer, 1998).

Clause-internally, all quantifiers can take existential and distributive scope over material they c-command. Whether inverse scope is possible at all is subject to cross-linguistic variation, and which quantifiers take inverse scope over which others depends on their syntactic hierarchy; therefore these matters are to some extent syntactic. But the classes of quantifiers that behave differently are semantically characterizable (Beghelli and Stowell, 1997; Szabolcsi, 1997b).

Both unmodified indefinites and definites pick out sets undergoing existential closure. As expected from the foregoing discussion, in (22) it is easy to keep the poems fixed (wide existential scope). But triplets of girls only marginally vary with poems, probably because predicates only marginally support inverse distributivity.

(22) Three girls know two poems/these poems/all the poems.

Universals like *every/each girl* readily take inverse distributive scope:

(23) Three girls know each/every poem.

Counting quantifiers may be intersective (*at least n ...*, *more/less than n ...*) or proportional (*few ...*).

They do not take inverse scope over universals (24) but they can make higher counters dependent (25):

(24) Every girl knows more/less than three/few poems.

(25) At least one girl knows more than three poems.

It is usually assumed that both universals and counters operate in the manner of generalized quantifiers, see (6b), and the distributivity of universals comes from lexical semantics. Hungarian presents evidence to the contrary. In that language, different types of quantifiers occupy different surface syntactic positions. Interestingly, the comparative quantifier *több, mint n* ... 'more than n ...' may occur either in the position of universals or in the position of counters. In the former case its interpretation is necessarily distributive; in the latter case it may be collective. This suggests that the former position has a distributive operator associated with it. *Every girl* is distributive because it only occurs in that position; its own contribution to interpretation is just the set of girls. This set plays a similar role as the one contributed by *two girls*. Counters may be the only quantifiers that truly operate in the manner of (6b). (Bare plurals on the existential reading are not regarded as quantifiers but as kind-denoters or as incorporated predicates.)

See also: Boole and Algebraic Semantics; Discourse Representation Theory; Dynamic Semantics; Formal Semantics; Game-Theoretical Semantics; Monotonicity and Generalized Quantifiers; Montague Semantics; Negation: Semantic Aspects; Plurality; Quantifiers: Semantics.

Bibliography

- Beghelli F & Stowell T (1997). 'Distributivity and negation.' In Szabolcsi (ed.), 71–108.
- Beghelli F, Ben-Shalom D & Szabolcsi A (1997). 'Variation, distributivity, and the illusion of branching.' In Szabolcsi (ed.), 29–70.
- Büring D (2004). 'Crossover situations.' *Natural Language Semantics* 12, 23–62.
- Cooper R (1983). *Quantification and syntactic theory*. Dordrecht: Reidel.
- Egli U & von Heusinger K (1995). 'The epsilon operator and E-type pronouns.' In Egli U *et al.* (eds.) *Lexical knowledge in the organization of language*. Amsterdam: John Benjamins. 121–141.
- Elbourne P (2001). 'E-Type anaphora as NP-deletion.' *Natural Language Semantics* 9, 241–288.
- Evans G (1980). 'Pronouns.' *Linguistic Inquiry* 11, 337–362.
- Fodor J D & Sag I (1982). 'Referential and quantificational indefinites.' *Linguistics and Philosophy* 5, 355–398.
- Groenendijk J & Stokhof M (1991). 'Dynamic predicate logic.' *Linguistics and Philosophy* 14, 39–100.
- Heim I (1982). *The semantics of definite and indefinite noun phrases*. Published (1989), New York: Garland.
- Heim I (1990). 'E-Type pronouns and donkey anaphora.' *Linguistics and Philosophy* 13, 137–177.
- Hendriks H (1993). *Studied flexibility: categories and types in syntax and semantics*. Amsterdam: Institute for Logic, Language and Computation, University of Amsterdam.
- Hintikka J (1997). 'No scope for scope?' *Linguistics and Philosophy* 20, 515–544.
- Hintikka J & Sandu G (1997). 'Game-theoretical semantics.' In van Benthem J & ter Meulen A (eds.) *Handbook of logic & language*. Cambridge, MA: MIT Press. 361–410.
- Jacobson P (1999). 'Towards a variable-free semantics.' *Linguistics and Philosophy* 22, 117–184.
- Kamp H & Reyle U (1993). *From discourse to logic*. Dordrecht: Kluwer.
- Kayne R (1998). 'Overt vs. covert movement.' *Syntax* 1, 128–191.
- Kratzer A (1998). 'Scope or pseudoscope? Are there wide scope indefinites?' In Rothstein S (ed.) *Events and grammar*. Dordrecht: Kluwer. 163–196.
- Kruijff G-J & Oehrle R (2003). *Resource-sensitivity, binding and anaphora*. Dordrecht: Kluwer.
- Landman F (2000). *Events and plurality: the Jerusalem lectures*. Dordrecht: Kluwer.
- May R (1985). *Logical form: its structure and derivation*. Cambridge, MA: MIT Press.
- Montague R (1974). *Formal philosophy: selected papers of Richard Montague*. Introduction by Thomason R (ed.). New Haven and London: Yale University Press.
- Neale S (1990). *Descriptions*. Cambridge, MA: MIT Press.
- Reinhart T (1997). 'Quantifier scope: How labor is divided between QR and choice functions.' *Linguistics and Philosophy* 20, 335–397.
- Schein B (1993). *Plurals and events*. Cambridge, MA: MIT Press.
- Szabolcsi A (1989). 'Bound variables in syntax (are there any?).' In Bartsch R, van Benthem J & van Emde Boas P (eds.) *Semantics and contextual expression*. Dordrecht: Foris. 295–318.
- Szabolcsi A (ed.) (1997a). *Ways of scope taking*. Dordrecht: Kluwer.
- Szabolcsi A (1997b). 'Strategies for scope taking.' In Szabolcsi (ed.), 109–154.

Semantic Value

J Dever, University of Texas, Austin, TX, USA

© 2006 Elsevier Ltd. All rights reserved.

A total theory of linguistic understanding is often taken to require three subtheories: a syntactic theory, a semantic theory, and a pragmatic theory. The semantic theory occupies an intermediary role – it takes as input structures generated by the syntax, assigns to those structures meanings, and then passes those meanings on to the pragmatics, which characterizes the conversational impact of those meanings. Semantic theories thus seek to explain phenomena such as truth conditions of and inferential relations among sentences/utterances, anaphoric relations among terms, and ambiguity and incoherence of expressions.

One way in which a semantic theory can provide the required explanations is by associating each expression provided by the syntax with a particular entity called its semantic value. These semantic values then serve both to ground the desired semantic explanations and to provide the pragmatic theory with input material on which to operate. While almost any approach to semantic theorizing can be reified into a theory of semantic value – a Davidsonian truth-theoretic account, for example, can associate atomic expressions with axioms and complex expressions with derivations; a translational theory can associate expressions with expressions in the target language of translation – setting semantic theory in the context of semantic value is most typical of an approach to semantics running through Frege, Carnap, and Montague. This discussion begins with the relatively simple theory of semantic values found in Frege, and progresses through various complications of that basic formula. First, the Fregean framework is extended to a full semantic type hierarchy. Second, the Fregean use of truth as the foundational semantic value is expanded to a notion of indexed truth as the foundational semantic value is expanded to a notion of indexed truth suitable for intensional and context-sensitive languages. Finally, dynamic semantic theories are considered, which diverge from traditional semantic theories by using qualities other than truth to construct semantic values.

Writing in 1891, Frege in ‘Function and concept’ considers the function $x^2 = 1$ formed by abstracting from an expression such as $(-1)^2 = 1$. Observing that the result of replacing x by various arguments is sometimes a true equation and sometimes a false equation, he says:

I now say: ‘the value of our function is a truth-value [*Wahrheitswert*],’ and distinguish between the

truth-values of what is true and what is false. I call the first, for short, the True; and the second, the False (Frege, 1997: 137).

Frege’s shift from the adjectives ‘true’ and ‘false’ to the neologicistic nominals ‘the True’ and ‘the False’ marks the beginning of semantic theorizing grounded in semantic values. ‘*Wahrheitswert*’ then crosses into English as ‘truth value’ in Russell’s commentary on Frege in 1903 in *The principles of mathematics*, and the contemporary ‘semantic value’ appears to derive from this usage.

Frege’s semantic theory associates with each expression in the language a referent (*Bedeutung*). The various referents are then the semantic values of the expressions in the language. Two central theses governing Frege’s conception of semantic value can be identified:

1. The two fundamental types of semantic values are truth values and objects (the referents of singular terms). Other semantic values are derived from the fundamental types through the construction of functions. Call this the categorical principle. Frege recognizes the derived categories of (a) first-level n -ary functions from objects and truth values to objects or truth values (paradigmatically, predicates and truth-functional connectives) and (b) second-order n -ary functions from first-order n -ary functions to objects or truth values (paradigmatically, quantifiers and definite descriptors).
2. Semantic values of complex expressions are derived from semantic values of their composite expressions, typically (although not necessarily) via functional application of one semantic value to another. Call this the compositional principle.

A standard semantic theory for quantified first-order logic can be formulated in Fregean terms by assigning to predicates the characteristic functions of their extensions, to connectives functions from truth values to truth values, and to quantifiers functions from predicate-assigned functions to truth values. The resulting theory will, as desired, account for truth conditions of and inferential connections between sentences. Thus, for example, semantic values can be assigned as follows:

- The predicate ‘ F ’ interpreted as *is French*, is assigned the function f from objects to truth values such that $f(x)$ is the True if and only if x is French.
- The connective ‘ \vee ,’ interpreted as *or*, is assigned the function g from pairs of truth value to truth values such that $g(x, y)$ is the True if and only if at least

one of x and y is the True. The connective ‘ \neg ,’ interpreted as *not*, is assigned the function h from truth values to truth values such that $h(x)$ is the True if and only if x is the False.

- The quantifier ‘ \exists ,’ interpreted as *something*, is assigned the function j such that $j(x)$ is the True if and only if x is a function from objects to truth values, which does not map every object to the False.

From these semantic values, it can be determined that $\exists xFx$ is true if and only if something is French, and that $\exists x(Fx \vee \neg Fx)$ is a logical truth.

The basic Fregean framework of semantic values coupled with the categorical and compositional principles then generalizes in three significant directions in later semantic work. First, the space of semantic values is extended to a full type hierarchy based on truth values and objects. The roots of this extension can be seen in Tarski’s (1933) paper ‘The concept of truth in formalized languages,’ but the full implementation of the thought in semantic theorizing about natural languages occurs with Montague’s ‘The proper treatment of quantification in ordinary English’ and Lewis’s ‘Generalized semantics.’ Introductory treatments can be found in Heim and Kratzer’s *Semantics and generative grammar* and in the Dowty *et al.*’s (1980) *Introduction to Montague semantics*.

In Montagovian semantics, a categorical grammar is linked with a type hierarchy of semantic values. The categorical grammar takes a small collection of primitive syntactic categories. Ajdukiewicz’s work introducing categorical grammars in ‘Syntactic connexion’ uses N (name) and S (sentence) as the primitive categories. Derived categories are then defined as follows:

- $\langle \alpha/\beta \rangle$ = the category of expressions combining with a β -category expression to form an α -category expression.

Thus:

- Intransitive verbs can be of category $\langle S/N \rangle$. ‘Snores’ can combine with an expression of category N (‘Socrates’) to form an expression of category S (‘Socrates snores’).
- Nouns can also be of category $\langle S/N \rangle$. ‘Linguist’ can combine with an expression of category N (‘Socrates’) to form an expression of category S (‘Socrates is a linguist’). Placement of nouns in $\langle S/N \rangle$ is somewhat syntactically forced and is influenced by the treatment of nouns as predicates in first-order logic; other versions of categorical grammar contain a third primitive category T of nouns.

- Modal operators can be of category $\langle S/S \rangle$. ‘Necessarily’ can combine with an expression of category S (‘Aristotle is fond of dogs’) to form an expression of category S (‘Necessarily, Aristotle is fond of dogs’).
- Quantified noun phrases can be of category $\langle S/\langle S/N \rangle \rangle$. ‘Some linguist’ can combine with an expression of category $\langle S/N \rangle$ (‘snores’) to form an expression of category S (‘Some linguist snores’).
- Determiners can be of category $\langle \langle S/\langle S/N \rangle \rangle / \langle S/N \rangle \rangle$. ‘Some’ can combine with an expression of category $\langle S/N \rangle$ (‘linguist’) to form an expression of category $\langle S/\langle S/N \rangle \rangle$, which in turn combines with an expression of category $\langle S/N \rangle$ (‘snores’) to form an expression of category S (‘Some linguist snores’).

The type hierarchy of semantic values similarly takes a small collection of primitive semantic categories. Following the Fregean tradition (and setting aside issues of intensionality to be raised below), the primitive categories can be t (the set of truth values) and e (the set of objects). Derived categories are then defined as follows:

$\langle \tau, \pi \rangle$ = the set of functions from category π to category τ .

If we then associate each syntactic category with a semantic category via a mapping $\llbracket \cdot \rrbracket$ such that:

1. $\llbracket N \rrbracket = e$
2. $\llbracket S \rrbracket = t$
3. $\llbracket \langle \alpha/\beta \rangle \rrbracket = \langle \llbracket \alpha \rrbracket / \llbracket \beta \rrbracket \rangle$

we can then obtain a semantic theory in which the semantic value of any complex expression results from the functional application of the semantic value of one of its immediate constituents to the semantic value of the other of its constituents. Both nouns and intransitive verbs, for example, are assigned functions from objects to truth values. Those functions can then be treated as characteristic functions determining predicate extensions.

This broadly Montagovian framework provides a fertile setting for much work in formal semantics. Barwise and Cooper’s (1980) work on generalized quantifiers, for example, fits naturally into such an approach. The categorical grammar yields the categories of $\langle \langle S/\langle S/N \rangle \rangle / \langle S/N \rangle \rangle$ for determiners and of $\langle S/\langle S/N \rangle \rangle$ for quantified noun phrases. The type hierarchy of semantic values then assigns values of type $\langle \langle t/\langle t/e \rangle \rangle / \langle t/e \rangle \rangle$ to determiners and of type $\langle t/\langle t/e \rangle \rangle$ to quantified noun phrases. ‘All philosophers’ for example, is assigned a function from extensions (in the form of

characteristic functions from objects to truth values) to truth values. The appropriate function is that which maps to the True sets that have the set of philosophers as a subset, and to the False other sets. The determiner 'all' then receives as semantic value a function from $\langle t/e \rangle$ values (predicate extensions) to quantified noun phrase values $\langle t/\langle t/e \rangle \rangle$. The appropriate function for 'all' is that which maps each set X of objects to the function that maps a set Y to the True if and only if $X \subseteq Y$. The type hierarchy yields a natural space of all possible quantifiers, allowing the location of semantic universals governing quantifiers in natural language, of the sort Barwise and Cooper set out. Thus, monotone increasing quantifiers, such as 'some linguist' and 'every mathematician,' are assigned functions f such that, for any extensions X and Y , if $f(X)$ is the True and $X \subseteq Y$, then $f(Y)$ is also the True. Monotone increasing quantifiers thus support inferences of the form:

Some linguist owns a red car. Therefore, some linguist owns a car.

in which the predicate of the conclusion is more restrictive than that of the premise. Monotone decreasing quantifiers, on the other hand, are assigned functions f such that, for any extensions X and Y , if $f(X)$ is the True and $Y \subseteq X$, then $f(Y)$ is also the True. Monotone decreasing quantifiers such as 'no archaeologist' and 'few physicists' support inferences of the form:

Few physicists own a red car. Therefore, few physicists own a car.

in which the predicate of the conclusion is less restrictive than that of the premise. Not all quantifiers are monotone in either direction, as examples like 'an even number of chemists' show – Barwise and Cooper thus hypothesize that all simple natural language quantifiers are conjunctions of monotone quantifiers.

Many complications of the simple type hierarchy of Montagovian semantics can now be investigated. Partee and Rooth's 'Generalized conjunction and type ambiguity,' for example, argues that expressions cannot be assigned a single stable semantic category. In the sentence:

John caught and ate a fish

the transitive verbs 'caught' and 'ate' need to be assigned the category $\langle \langle t/e \rangle / e \rangle$. If each transitive verb is assigned a function from entities (e) to intransitive verb extensions ($\langle t/e \rangle$), then their conjunction will be assigned the function that maps any entity x to the intransitive verb that assigns the True to any

entity that caught and ate x . This function can then combine with the entity assigned to 'John' to create an appropriate truth value for the whole sentence. However, this category assignment for transitive verbs fails in other examples such as:

John needed and bought a new coat.

The truth value of this sentence does not depend on a function that maps entities x to intransitive verbs assigning the True to entities that needed and bought x , since we do not here require that John needed and bought any one particular entity. Instead, we require that John needed and bought one type of entity. Thus, the transitive verbs 'needed' and 'caught' must be assigned the category $\langle \langle t/e \rangle / \langle t/\langle t/e \rangle \rangle \rangle$. By assigning to each transitive verb a function from nouns ($\langle t/e \rangle$) to truth values, their conjunction will then be assigned the function that assigns the True to any object that needed and bought that type of object. Partee and Rooth (2002) thus suggest that the semantic value assigned to an expression can undergo type raising, which moves it from a simpler to a more complex type to fit constructions such as these.

The second generalization of the core Fregean framework is the move to truth at a specification point. The first instance of this is the intensionalization of Frege's extensional logic. The roots of intensionalization can already be seen in Frege's separation of sense and reference in 'On sense and reference,' but the full picture emerges in Carnap's *Meaning and necessity*. Carnap introduces state descriptions – negation-complete sets of atomic sentences equivalent to possible worlds – and defines truth in a state description. He then assigns to every expression both an extension and an intension. Its extension is its semantic value in the Fregean truth value/object-based type hierarchy, and its intension is its semantic value in a parallel hierarchy based on functions from state descriptions to truth values and functions from state descriptions to objects. Given a designation of an extensionally privileged state description (the actual world), the extensional type hierarchy can be subsumed into the intensional. The intensional type hierarchy is connected conceptually to the Fregean type hierarchy via an understanding of truth at a world in terms of truth simpliciter (what would be true, were the world a certain way). This conceptual connection allows the intensional semantics still to address questions of truth conditions and inferential connections.

The basic insight of the intensionalizing move is that a finer-grained semantics can be achieved by replacing truth values simpliciter in the type hierarchy

with indexed truth values, in the form of functions from some set of indexes to truth values. The result is a semantic theory equipped for nontruth-functional operators such as modalities. A similar approach can thus accommodate tense by moving to truth at a time (or at a time/world pair). The same insight is employed in adapting the Fregean framework to context-sensitive expressions. Kaplan, in 'Demonstratives,' gives a semantic theory in which expressions are assigned intensions (both modal and temporal) with respect to a context, where a context provides information at least about utterance speaker, spatial and temporal location, and world. Kaplan's system thus uses a form of double-indexing; the background type hierarchy is grounded on functions from pairs of indices $\langle\langle\text{world}, \text{time}\rangle, \text{context}\rangle$ to truth values and objects. Thus, the sentence:

I am a philosopher

can be assigned a truth value only relative first to a context that determines the referent of the indexical 'I' and second to a world-time pair that determines the properties of the contextually provided referent. Kamp's approach to tense in 'Formal properties of now' uses a similar system of double indexing, and Stalnaker (1978) in 'Assertion' gives a streamlined version of the double-indexing approach, based on Segerberg's (1973) two-dimensional modal semantics, in which both indices are simply worlds.

Context-sensitive semantic theories thus extend the Fregean framework further in the direction begun by intensional semantics, by extending the indexing of truth at the foundation of the type hierarchy. The introduction of double-indexing gives rise to a novel issue in interpretation. The single-indexing of intensional semantics typically comes with a collection, in the analyzed language, of intentional operators that act on the indexing position; thus the expression 'necessarily' binds the world index in the semantic value of expressions it governs, causing them to be evaluated at every possible world. As a consequence, the new foundational semantic category of truth-at-an-index can be reductively understood via the simple truth of some claim containing an intensional operator. For example, a claim ϕ 's being true at a world w can be reduced to the intensionally modified claim 'If it were the case that w^* , it would be the case that ϕ 's being simply true' (where w^* is a canonical description of world w). However, in double-indexed semantic theories, the second indexed position is typically not subject to semantic interaction with intensional operators of the language. In a sentence such as:

I am always here.

the intensional operator 'always' controls the first index, governing the temporal evaluation of the verb. It does not, however, govern the second index, determining the contextual assignment of semantic values to 'I' and 'here.' Double-indexing thus means that truth at an index can no longer be understood in terms of what would be (or was, or will be) a true simpliciter. It requires instead a novel understanding of the elements of the type hierarchy. The result is the idea that semantic value characterized at a certain level ('utterance meaning') varies based on context of use, so that the same expression can have different semantic contents in different contexts of use.

The third and final generalization of the Fregean framework is the move to dynamic semantics. Dynamic semantics can be thought of as the other side of the coin of context sensitivity, investigating the way that expressions affect, as well as are affected by, context. Dynamic semantic theories have flourished over the last 25 years, taking such forms as discourse representation theory, file change semantics, dynamic predicate logic, and update semantics. When placed in a compositional form, one distinctive characteristic of dynamic semantic theories is a move away from indexed truth to a nontruth-based fundamental semantic value. Overviews of major topics in dynamic semantics can be found in Gamut's (1991) *Logic, language, and meaning* and in Kamp and Reyle's (1993) *From discourse to logic*.

Groenindijk and Stokhof's (1991) dynamic predicate logic, for example, makes dynamic the semantic analysis of sentences of first-order logic using sets of satisfying assignments to variables (where truth is equated with satisfaction by all assignments). In dynamic logic, a sentence is associated not with a static set of assignments, but with a function from input assignments to output assignments. An existential claim of the form ' $\exists x Fx$ ' pairs an incoming assignment g with an outgoing assignment h if and only if h satisfies Fx and there is some assignment $g \in G$ such that g and h differ at most in the x position. Quantifiers thus have unrestricted rightward scope; by passing on assignments whose x positions satisfy the existentially quantified matrix, future reference to and description of these objects is then possible. This unrestricted rightward scope allows dynamic predicate logic to model the behavior of crossclausal anaphora on indefinite noun phrases, as in:

A man walked in the park. He wore a hat.

The dynamic move to nontruth-based semantic values is more radical than the intensionalizing

move to semantic values based on truth at an index. One consequence is that associated semantic concepts require redefinition. Equivalence, for example, can no longer be understood as sameness of truth value in all models, in the absence of truth-based semantic values. Dynamic predicate logic allows multiple notions of equivalence, the stronger of which is identity of behavior in input–output conditions:

$\phi \simeq \psi$ if ϕ pairs an input assignment g with an output assignment h if and only if ψ does as well.

$\exists xFx \wedge Gx$ is not equivalent to $Gx \wedge \exists xFx$ in this sense, since an assignment g whose x position does not satisfy Gx will produce no output when input to $Gx \wedge \exists xFx$, but can produce an output when input to $\exists xFx \wedge Gx$, since in the latter case the x value can be ‘reset’ to an object satisfying F and G , if there is such an object. Similarly, dynamic predicate introduces a new consequence relation:

$\phi \models \psi$ if and only if for any model \mathcal{M} and any assignment g in \mathcal{M} , if the input of g to ϕ produces some output assignment h , then the input of h to ψ produces some further output assignment k .

The resulting consequence relation differs from classical consequence relations in many structural features. Idempotence fails; we do not have $Gx \wedge \exists xFx \models Gx \wedge \exists xFx$, since the output assignment, with its reset value in the x position, need not satisfy Gx . Transitivity similarly fails, since $\neg\neg\exists xFx \models \exists xFx$ and $\exists xFx \models Fx$, but $\neg\neg\exists xFx \not\models Fx$, since negation blocks dynamic effects.

See also: Character versus Content; Compositionality: Philosophical Aspects; Context Principle; Discourse Representation Theory; Dynamic Semantics; Extensionality and Intensionality; Meaning: Overview of Philosophical Theories; Quantifiers: Semantics; Semantics–Pragmatics Boundary; Syntax–Semantics Interface

Bibliography

- Ajdukiewicz K (1967). ‘Syntactic connexion.’ In McCall S (ed.) *Polish logic*. Oxford: Clarendon Press.
- Barwise J & Cooper R (1980). ‘Generalized quantifiers and natural language.’ *Linguistics and Philosophy* 4, 159–219.
- Carnap R (1947). *Meaning and necessity*. Chicago: The University of Chicago Press.
- Davidson D (1984). ‘Truth and meaning.’ In *Inquiries into truth and interpretation*. Oxford: Oxford University Press.
- Dowty D, Wall R & Peters S (1980). *Introduction to Montague semantics*. Kluwer Academic Publishers.
- Frege G (1997a). ‘Function and concept.’ In Beaney M (ed.) *The Frege reader*. Blackwell Publishers.
- Frege G (1997b). ‘On sinn and bedeutung.’ In Beaney M (ed.) *The Frege reader*. Blackwell Publishers.
- Gamut L T F (1991). *Logic, language, and meaning*. Chicago: University of Chicago Press.
- Groenendijk J & Stokhof M (1991). ‘Dynamic predicator logic.’ *Linguistics and Philosophy* 14, 39–100.
- Heim I (1983). ‘File change semantics and the familiarity theory of definiteness.’ In Bauerle R, Schwarze C & von Stechow A (eds.) *Meaning, use and the interpretation of language*. Walter de Gruyter & Co.
- Heim I & Kratzer A (1998). *Semantics in generative grammar*. Blackwell.
- Janssen T (1997). ‘Compositionality.’ In van Benthem J & ter Meulen A (eds.) *Handbook of logic and language*. The MIT Press.
- Kamp H (1970). ‘Formal properties of “now.”’ *Theoria* 37, 227–273.
- Kamp H (1984). ‘A theory of truth and semantic representation.’ In Groenendijk J, Janssen T & Stokhof M (eds.) *Truth, interpretation, and information: selected papers from the third Amsterdam colloquium*. Foris.
- Kamp H & Reyle U (1993). *From discourse to logic: introduction to model-theoretic semantics of natural language*. Kluwer Academic Publishers.
- Kaplan D (1989). ‘Demonstratives.’ In Almog J, Perry J & Wettstein H (eds.) *Themes from Kaplan*. Oxford: Oxford University Press.
- Lewis D (1970). ‘General semantics.’ *Synthese* 22, 18–67.
- Montague R (2002). ‘The proper treatment of quantification in ordinary English.’ In Portner P & Partee B (eds.) *Formal semantics: the essential readings*. Blackwell Publishers.
- Partee B & Rooth M (2002). ‘Generalized conjunction and type ambiguity.’ In Portner P & Partee B (eds.) *Formal semantics: the essential readings*. Blackwell Publishers.
- Quine W V O (1960). *Word and object*. The MIT Press.
- Russell B (1903). *The principles of mathematics*. W. W. Norton & Company.
- Segerberg K (1973). ‘Two-dimensional modal logic.’ *Journal of Philosophical Logic* 2, 77–96.
- Stalnaker R (1978). ‘Assertion.’ In Cole P (ed.) *Syntax and semantics* 9. Academic Press.
- Tarski A (1983). ‘The concept of truth in formalized languages.’ In *Logic, semantics, metamathematics*. Hackett Publishing Company.
- Veltman F (1996). ‘Defaults in update semantics.’ *Journal of Philosophical Logic* 25, 221–261.

Semantics of Interrogatives

B Sherman, Princeton University, Princeton, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

Many problems in semantics are typically expressed as problems about how to define a function from the form of a sentence to the conditions in which the sentence can be uttered truly (or, alternatively, to a proposition that can be expressed by an utterance of the sentence). The main problem posed by interrogatives for the truth-conditional approach to semantics is rather different. They do not give rise to a problem of how to map a function from the form of a sentence to its truth conditions. Rather, they give us reason to suspect that there is no such function. For interrogatives, unlike indicatives, do not seem to have truth conditions, and nor do they seem to express propositions that themselves have truth conditions. The main problem interrogatives pose is metasemantic: to say what a semantics for interrogatives must do.

In the 'Metasemantics' section of this article, I survey three attempts to deal with this question. In the 'Semantics' section, I present some basic-level semantic questions that would need to be addressed even if the metasemantic question were settled – though in all probability they will in fact **help** us to settle the metasemantic question.

Metasemantics

Above, I introduced one constraint on choosing a semantic framework for interrogatives: interrogative sentences are not true or false, so we should not assign to them some semantic object that is true or false, at least not without further explanation. Another is the Davidsonian constraint of 'semantic innocence': the fixed meaning of a word should not vary with the environment in which it occurs. This constraint generates a tension with any approach to the metasemantic problem that ignores the truth-conditional approach that works so nicely for indicative sentences. For this reason, the three dominant frameworks that I will focus on all treat the semantics of interrogatives in broadly truth-conditional terms. That is, in all of the frameworks to be discussed here, the meaning of an interrogative is in some way tied to a propositional object of some sort. In the rest of this section, I will look at the following three approaches to the metasemantic problem: the force/radical approach, the epistemic-imperative approach, and the question-as-answer approach.

According to the force/radical approach, sentences can be factored into two components: a force indicator and a propositional radical. Roughly, the

propositional radical supplies the truth-conditional content to the sentence, and the force indicator suggests the attitude taken toward that content. So, on this approach, advanced by McGinn (1977), Davidson (1984), and Stainton (1999), sentences (1) and (2) have the same radical and different moods:

- (1) The monkey is hungry.
- (2) Is the monkey hungry?

The mood of the sentence indicates the force of a typical utterance of the sentence. Sentences in the indicative mood are typically used for asserting the proposition expressed by the radical. Sentences in the interrogative mood are typically used to ask whether the proposition expressed by the radical is true.

This picture works well for yes/no questions, such as (2). What about wh-questions, such as (3)?

- (3) What does the monkey want?

The radical of (3) could be thought of as a propositional function of the form in (4):

- (4) λx : the monkey wants x

Some story would then be needed to explain why attaching an interrogative mood to (4) would yield the desired interpretation. In addition, it seems like the only mood one could attach to (4) is the interrogative mood. This seems out of character with the general approach, which emphasizes the possibility of attaching different moods to the same radical, thus requiring some explanation if the approach is to be defended.

The epistemic-imperative approach treats the meaning of an interrogative as an imperative concerning one's epistemic state. Sentence (2) would be analyzed as meaning what (5) means, and (3) would be analyzed as meaning what (6) means.

- (5) Let it be the case that (or: Bring it about that)
I know whether the monkey is hungry.
- (6) Let it be the case that (or: Bring it about that)
I know what the monkey wants.

On this approach, advocated by Åqvist (1965) and Hintikka (1983), the direct interrogative exemplified by (2) and (3) is analyzed in terms of the indirect interrogative contained in (5) and (6). In order for this approach to provide a semantic framework for interrogatives, an account of the semantics of indirect interrogatives is needed.

Hintikka gave an analysis of 'know whether' in terms of 'know that.' Sentence (5) on this analysis would be equivalent to (7):

- (7) Bring it about that either I know that the monkey is hungry, or I know that the monkey is not hungry.

However, this analysis of indirect interrogatives faces difficulties when embedded under verbs such as 'wonder':

- (8) I wonder whether the monkey is hungry.

Sentence (8) does not seem susceptible to a similar sort of paraphrase, causing trouble for any view that attempts to analyze away the indirect interrogative.

Finally, the question-as-answer approach treats the interrogative, either direct or indirect, as denoting a question, where this is understood to be a special sort of semantic object. Since a question determines a set of answers, the set of answers is used as a surrogate object for the question (much in the way that a set of possible worlds is used as a surrogate object for propositions). The intuitive idea motivating this approach, originating in Hamblin (1958), is that to know the meaning of a question is to know what counts as an answer to it. Different versions of the approach differ in which set of answers is used.

On the most standard version of this approach, advanced by Hamblin (1958) and Groenendijk and Stokhof (1994), the meaning of an interrogative should be thought of as the set of possible answers to a question, where this set of possible answers forms a partition of logical space. Every possible answer on this view is a complete answer to the question, and the set of answers jointly exhausts the possibilities of complete answers.

Karttunen (1977) provided two reasons for treating the meaning of an interrogative as the set of its true, rather than possible, answers. First, consider sentence (9):

- (9) Who is elected depends on who is running.

According to Karttunen, this sentence says that the true answer to the subject position question depends on the true answer to the object position question. So treating the meanings of interrogatives as sets of true answers provides a more straightforward account of verbs such as 'depend on.' Second, consider sentences (10) and (11):

- (10) John told Mary that Bill and Susan passed the test.

- (11) John told Mary who passed the test.

Sentence (11) entails that John told Mary the truth, whereas this is not the case with sentence (10). By treating the indirect interrogative in (11) as denoting a set of true answers, this entailment is straightforwardly explained.

Karttunen's account has some counterintuitive consequences, however. For example, somebody who asks question (2) in a situation where the monkey is hungry intuitively asks the same question as somebody who asks question (2) in a situation where the monkey is not hungry. But on Karttunen's account, the meaning of the questions asked in the two situations is different, since the true answers to the questions asked are different.

A more general worry with the question-as-answer approach, articulated by Stainton (1999), is that it makes the domain of the interrogative – the set of objects that figure in the possible answers to it – a part of its meaning. It seems intuitive that one can understand a question without knowing anything about the objects that figure in the possible answers to it. For example, if an alien from outer space lands on Earth, we might ask the question 'What does the alien want?' Surely, among the possible answers to this question are objects that we've never seen or imagined before. But that doesn't stop us from understanding the meaning of the question, as it seems it should on the question-as-answer approach.

Although the approaches to the metasemantic problem presented here are dominant in the literature, they are by no means exhaustive. Ultimately, the success of a given framework will depend on the extent to which it is successful in accounting for various semantic phenomena.

Semantics

One of the ongoing semantic debates in the semantics of interrogatives concerns the ambiguity that results from a *wh*-question containing a universal quantifier, as in (12):

- (12) Who does everyone like?

According to one reading of (12), the question is asking which people are such that every person likes them. An appropriate answer might be, for example, 'Bill and Mary.' On the other reading of (12), the question is asking which person each person likes. An appropriate answer to this question would be a list of pairs of people of the form 'Bill likes Mary, Mary likes Sue, Sue likes Bill.' The debate specifically concerns this second reading of the question, called the *pair/list* reading. The question is how to account for that reading semantically.

There have been two main sides to the debate. According to one standard view, the role of the quantifier that occurs in the interrogative is to restrict the domain of the question. So (12) could be paraphrased roughly as (13):

- (13) For every person, who does that person like?

Exactly how this paraphrased reading is derived from (12) will depend in large part on the metasemantic approach that one favors. For example, Groenendijk and Stokhof (1984) modified Karttunen's metasemantic approach in implementing their view.

According to an alternative view, pair/list readings are an instance of a more general kind of reading. Consider the following question/answer pair in (14):

- (14a) Who does every man love?

- (14b) His mother.

The answer in (14) gives rise to a functional reading of the question. The answer to the question is not an individual, but a function – that is, a rule which takes one from an object to another object. Pair/list readings are, according to this alternative, functional readings, where the function is specified extensionally, in terms of the ordered pairs in the extension of the function. This view was presented in Engdahl (1985) and developed by Chierchia (1993). Accounting for functional readings is an interesting semantic issue in its own right. For alternative accounts of pair/list readings, see Beghelli (1997), Szabolcsi (1997a), and Pafel (1999).

Another semantic issue concerns the nature of the presuppositions that different sorts of questions give rise to. For example, the question 'What is it like owning a monkey?' presupposes that the addressee owns a monkey. The question 'Who came to the party?' presupposes that someone came to the party. The question 'Which monkey ate the banana?' presupposes that a unique monkey ate the banana. Whether these presuppositions are semantic in nature and, if so, where they arise from has been a contested issue. See Belnap and Steel (1976), Karttunen (1977), Higginbotham and May (1981), and Hintikka (1983).

A final semantic issue that is of both linguistic and philosophical interest concerns the context sensitivity involved in whether something counts as an answer to a question. This issue is particularly pressing for the question-as-answer approach, since most versions of that approach assume that each question has some unique complete answer. Ginzburg (1995) developed a novel account of the semantics of interrogatives aimed at accommodating various sorts of context sensitivity. In general, questions calling for the identification of something seem to be interest relative. For example, the sentence 'Where am I?' might be used to ask for the country in which one is located, the street on which one is located, the room in which one is located, etc. In some situations, one counts as knowing who killed Lady Chittlesworth when one knows

simply that the murderer is the person that was wearing the yellow shirt. In other situations, this would not count as an acceptable answer. See Boër and Lycan (1985) for an account of knowing who someone is.

This brief survey of issues is far from exhaustive. For a very useful overview of both metasemantic approaches to interrogatives and semantic issues concerning interrogatives, see Groenendijk and Stokhof (1994).

See also: Mood, Clause Types, and Illocutionary Force; Truth Conditional Semantics and Meaning.

Bibliography

- Åqvist L (1965). *A new approach to the logical theory of interrogatives 1: Analysis*. Uppsala: Uppsala Universiteit.
- Beghelli F (1997). 'The syntax of distributivity and pair-list readings.' In Szabolcsi (ed.). 349–408.
- Belnap N & Steel T (1976). *The logic of questions and answers*. New Haven: Yale University Press.
- Boër S & Lycan W (1985). *Knowing who*. Cambridge, MA: MIT Press.
- Chierchia G (1993). 'Questions with quantifiers.' *Natural Language Semantics* 1(2), 181–234.
- Davidson D ([1979] 1984). 'Moods and performances.' In *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
- Engdahl E (1985). *Constituent questions*. Dordrecht: Reidel.
- Ginzburg J (1995). 'Resolving questions I & II.' *Linguistics and Philosophy* 18, 459–527, 567–609.
- Groenendijk J & Stokhof M (1984). *Studies on the semantics of questions and the pragmatics of answers*. Ph.D. thesis, University of Amsterdam.
- Groenendijk J & Stokhof M (1994). 'Questions.' In van Benthem J & ter Meulen A (eds.) *Handbook of logic and language*. Amsterdam: Elsevier.
- Hamblin C (1958). 'Questions.' *Australasian Journal of Philosophy* 36, 159–168.
- Higginbotham J & May R (1981). 'Questions, quantifiers, and crossing.' *Linguistic Review* 1, 41–79.
- Hintikka J (1983). 'New foundations for a theory of questions and answers.' In Kiefer F (ed.) *Questions and answers*. Dordrecht: Reidel. 159–190.
- Karttunen L (1977). 'Syntax and semantics of questions.' *Linguistics and Philosophy* 1, 3–44.
- McGinn C (1977). 'Semantics for nonindicative sentences.' *Philosophical Studies* 32, 301–311.
- Pafel J (1999). 'Interrogative quantifiers within scope.' *Linguistics and Philosophy* 22, 255–310.
- Stainton R (1999). 'Interrogatives and sets of answers.' *Critica* 91, 75–90.
- Szabolcsi A (1997a). 'Quantifiers in pair-list readings.' In Szabolcsi A (ed.). 311–348.
- Szabolcsi A (ed.) (1997b). *Ways of scope taking*. Dordrecht: Kluwer Academic.

Semantics–Pragmatics Boundary

A Bezuidenhout, University of South Carolina,
Columbia, SC, USA

© 2006 Elsevier Ltd. All rights reserved.

The Philosophical Debate

Texts in the philosophy of language frequently cite the tripartite distinction between syntax, semantics, and pragmatics made by Morris (1938). According to Morris, syntax is concerned with the structural properties of signs (i.e., with word-word relations), semantics with the relations between signs and the things they signify (i.e., with word-world relations), and pragmatics with the uses of signs by speakers and hearers to perform communicative acts (i.e., with word-user relations).

Philosophers generally follow Frege in rejecting any form of mentalist semantics. They think of languages as “abstract semantic systems whereby symbols are associated with aspects of the world” (Lewis, 1972: 170). There is a potential infinity of both syntactically well-formed (grammatical) and semantically well-formed (meaningful) sentences in any language, and it is the job of semantics to identify rules that generate this potential infinity. On the other hand, since we are interested in the semantics of **natural** languages, these rules must be ones that are learnable by humans with finite minds. Semanticists are interested in what a competent speaker knows when she knows a language (i.e., her syntactic and semantic competence). Hence they assume that what a speaker knows is a **finite** set of rules that can compositionally generate the potential infinity of syntactically well-formed sentences and that can deliver a **semantic interpretation** for every meaningful sentence of the language.

Philosophers generally assume that there is a sharp division between syntax/semantics and pragmatics. While semantics studies the rules that a competent speaker knows when she knows the meanings of sentences, pragmatics studies how sentences are used in conversational contexts to communicate a speaker’s messages. Pragmatics is thus concerned with linguistic **performance** rather than competence. It is **by** using sentences with certain syntactic and semantic properties that speakers succeed in communicating certain things. So, the central question of pragmatics is how we succeed in our communicative tasks.

Many philosophers are convinced that Grice (1975, 1989) made an important start in answering this question by articulating his Cooperative Principle and maxims of conversation. Grice sees conversations as rational cooperative activities where hearers use

their linguistic knowledge, together with mutually available nonlinguistic contextual knowledge, to infer what the speaker means to communicate. The principles that guide conversations are analogous to the principles that guide any sort of rational cooperative activity, such as the joint activity of building a house or sailing a ship. Pragmatic principles on this view are not tied essentially to any language mechanism and are certainly not language-specific rules, unlike the syntactic and semantic rules that define a language.

An alternative view, argued for by Prince (1988, 1997), assumes that there are rules of use associating certain linguistic forms with certain functions. Moreover, these rules are language-specific, in the sense that the same pragmatic function could be served in different languages by different forms; so any competent speaker of the language must learn these rules. Knowledge of these rules constitutes the speaker’s pragmatic competence. Hence it is incorrect to put the study of pragmatics on the performance side of the competence/performance divide. Related to Prince’s ideas are those of Kasher (1991), who argues for a modular conception of pragmatics. Just as linguists have postulated a grammar module, so Kasher argues there is a module governing pragmatic processes, with its own proprietary rules and representations. Since it is Grice’s conception of pragmatics that has set the agenda for debate in the philosophy of language, these alternative views will be set aside here.

Gricean pragmatics introduces the idea that it is **by** saying certain things in certain contexts that speakers are able indirectly to communicate (to implicate) certain further things. In working out what a speaker has implicated, a hearer will use his knowledge of the conversational maxims, together with contextually available knowledge, to infer what the speaker communicated. For example, after a terrible ordeal in which a man is rescued from a remote mountainside after a plane crash, a TV reporter interviews him. The reporter asks: ‘Were you ever afraid?’ and the man replies: ‘I felt a twinge or two.’ By his understatement he has implicated that things were pretty bad. The understatement is a violation of Grice’s first Maxim of Quantity, which enjoins speakers to say as much as is required by the purposes of the talk exchange. The hearer, having recognized the violation, but assuming that the speaker is still bound by the Cooperative Principle, will search the context for further information that the speaker might have intended to convey. It is his background knowledge – of human psychology, of the probable consequences of plane wrecks, and of the low probability of survivors being found

in remote, sparsely populated places – that allows the reporter to infer the speaker's intended meaning.

The currently dominant view in philosophy of language is that a theory of meaning for a language specifies the truth-conditions for each of the sentences of a language. It does this by specifying a finite set of rules that compositionally generates these truth-conditions. This truth-conditional approach to semantics has been grafted onto a Gricean view of pragmatics. It is generally accepted that saying and implicating are neatly separated. Saying is tied to sentence meaning and the expression of truth-conditional content. What a speaker says when she utters a sentence (i.e., the locutionary content of her utterance) corresponds to the truth-conditional content of the sentence. (Note that the notion of **saying** is not to be conflated with the notion of **stating**. The former is a locutionary act, namely the act of expressing some content. The latter is an illocutionary act. It is the expressing of some content with a particular illocutionary force.) Implicating is tied to (indirectly or implicitly) communicated content that can be inferred once the hearer has figured out what the speaker has (directly or explicitly) said. Since truth-conditional content is the province of semantics and implicature is the province of pragmatics, the saying/implicating divide goes along with a neat divide between semantics and pragmatics. Consequently, many believe that truth-conditions can be specified in a way that is essentially free from pragmatic considerations.

But there are problems with this Gricean view. One of the first indications of trouble for this view came from some observations by Cohen (1971). Others, such as Carston (1988, 2002), Levinson (1995, 2000) and Recanati (1989, 2004) have used examples similar to Cohen's to challenge the Gricean picture. Consider examples such as the following:

- (1) Mary fell pregnant and she got married.
- (2) Mary got married and she fell pregnant.

Grice would say that (1) and (2) have the same truth-conditional content, but that they implicate different things. (1) implicates (in a generalized way) that the pregnancy occurred before the marriage, whereas (2) implicates the opposite. (1) also implicates that the reason for Mary's marriage was her pregnancy. However consider examples such as the following:

- (3) If Mary fell pregnant and she got married, her grandma will be shocked.
- (4) If Mary got married and she fell pregnant, her grandma will be shocked.

According to Grice, the antecedents of the two conditionals have the same truth-conditional content.

Therefore, (3) and (4) should themselves have the same truth-conditional content, yet intuitively they do not. (3) could be true while (4) is false. It looks as though the implicated content of (1) and (2) has become incorporated into the truth-conditional content of (3) and (4). In other words, (3) and (4) in effect express the following:

- (3') If Mary fell pregnant *and then for that reason* she got married, her grandma will be shocked.
- (4') If Mary got married *and then* she fell pregnant, her grandma will be shocked.

Clearly, (3') and (4') differ in content, so it is not a problem if one is true and the other is false.

These appear to be cases of pragmatic intrusion into truth-conditional content. Such pragmatic intrusion creates a problem for Grice that Levinson (2000) calls 'Grice's Circle.' The trouble is that to figure out what is conversationally implicated, the hearer must first determine what is said (since it is **by** saying such-and-such that a speaker succeeds in implicating something else). However, in figuring out what was said by (3) or (4), it looks as though one must first determine their implicated contents.

Levinson (1995, 2000) argues that pragmatic intrusion is not problematic, since it is limited to the intrusion of **generalized** conversational implicatures (GCIs), and these, he argues, are **default meanings** that will be automatically triggered by the use of certain kinds of expressions. The derivation of GCIs is governed by various heuristic principles. For example, the I-Principle can be summarized in the slogan 'What is simply described is stereotypically and specifically exemplified.' It applies only to 'unmarked, minimal expressions' (Levinson, 1995: 97). It enjoins speakers to minimize what they say when their hearers are able to use contextually accessible information to enrich the informational content of their utterances. Conversely, it enjoins hearers to amplify or enrich the informational content of the speaker's utterance up to the point that they judge is the speaker's intended meaning. Since 'and' is the sort of minimal, unmarked expression that calls for a stereotypical interpretation, conjunctions such as (1) and (2) will be given an interpretation according to which the events described by the two conjuncts are temporally ordered. The net effect is that (3) and (4) will be understood to express (3') and (4') respectively.

Not everyone would agree that in (3) and (4) we have pragmatic intrusion into truth-conditional content. Cohen (1971) appeals to examples of embedded conjunctions that seem to affect the truth-conditions of the larger sentences in which they are embedded to argue for a **semantic ambiguity** account of 'and.' If the suggestion of temporal ordering associated with

(1) affects the truth-conditions of (3), Cohen concludes that this feature must be part of the semantically encoded meaning of ‘and.’ Since relations other than temporal ordering can be suggested by a conjunction, this view is committed to a multiple ambiguity account of ‘and.’ In addition to conjunctions such as (1) and (2), consider examples such as:

- (5) It is summer in Europe and winter in Australia.
- (6) The fan turned on and {as a result} a cool breeze blew through the room.
- (7) Peter took a shower and {while in the shower} he practiced his singing.

In some cases, as in example (5), ‘and’ expresses simple truth-functional ‘and,’ and the conjuncts can be reversed without changing the meaning. In others, such as (6), it expresses a causal relation ‘and as a result,’ or, as in (7), a temporal containment relation ‘and while.’ In these last two cases, the relations are asymmetric, and reversing the conjuncts changes the meaning. For instance, reversing (6) suggests a different causal scenario, where the breeze somehow turns on the fan. Examples could be multiplied, and for each case where a different relation is suggested, Cohen would have to posit yet another meaning for ‘and.’

Carston (1988) and Recanati (1989) argue against positing a semantic ambiguity for ‘and,’ maintaining instead that the contents represented between brackets in the above examples are **pragmatically determined aspects of what is said**. (See entry on **Pragmatic Determinants of What Is Said**.) Rather than being semantically ambiguous, ‘and’ is **semantically underspecified**. It will be pragmatically enriched in different ways, depending on the assumptions that are operative in the conversational context.

Carston and Recanati agree with Levinson that there is pragmatic intrusion. However, they point to embedded contexts, like the conditionals (3) and (4), to argue that the pragmatic content associated with (1) and (2) belongs to what is said, rather than being conversationally implicated. (Carston (2002) prefers to use Sperber and Wilson’s (1986) technical term ‘explicature’ instead of the term ‘what is said,’ since the latter has a commonsense usage that interferes with attempts at terminological regimentation.) If the pragmatic content of a simple sentence has an effect on the truth-conditional content of the compound sentences in which it is embedded, then that pragmatic content is part of what is said by the simple sentence, not something that is merely implicated. Recanati calls this the Scope Test.

Other tests have been proposed for determining whether some pragmatically determined content is part of what is said. Recanati (1989) proposes his

Availability Principle, according to which any content that intuitively seems to affect truth-conditions should be regarded as a part of what is said. Carston (1988) proposes her Functional Independence Principle, which requires both explicatures and implicatures to occupy independent roles in inferential interactions with other assumptions. Take example (8) discussed below. The simple encoded content that Mary engaged in an act of swallowing is not functionally independent of the enriched content that Mary swallowed a bug. The latter entails the former, and it is from the latter that further contextual effects can be derived. This suggests that it is the enriched content that corresponds to what is said, not the more minimal encoded content, which has no autonomous role to play.

Explicatures are pragmatic developments of semantically encoded content and can be either **enrichments** or **loosenings** of encoded content. Carston argues that the processes involved in the recovery of explicatures are inferential processes and, hence, no different from the sorts of inferential processes involved in the derivation of conversational implicatures. What distinguish explicatures from implicatures are not the sorts of processes involved in their derivation but the starting points of these inferential processes. Derivations of explicatures begin with the semantically underspecified representations of logical form (LF) that are the output of processes of grammatical decoding. Implicatures, on the other hand, as Grice insisted, are contextual implications that follow from contexts including assumptions about what was said (i.e., including explicatures). This is **not** a commitment to the claim that explicatures are processed before implicatures. In fact, Carston thinks that the processing of explicatures and implicatures happens in parallel, and that the overall interpretation of a speaker’s utterance is something arrived at via a process of mutual adjustment.

Recanati, on the other hand, distinguishes local from global pragmatic processes. The sorts of processes involved in the derivation of pragmatic determinants of what is said are of the local sort and are noninferential. For instance, such local processes might involve spreading activation within an associative conceptual network, or the accessing of stereotypical information from conceptual frames or scripts. Such processing happens at a subconscious level, and only the output of such processes is consciously available. In contrast, global pragmatic processes are inferential processes of the sort that Grice claimed are involved in the derivation of conversational implicatures. Such inferential processing is in principle consciously available, in the sense that language users can become aware not just of the conclusions of such

reasoning but also of the inputs to such reasoning, as well as to the (putative) fact that premises and conclusions are inferentially connected.

Bach (1994) provides yet another perspective on these matters. Bach wishes to maintain a more minimalist conception of semantics and of what is said. Yet he acknowledges that there are pragmatically determined contents that are not Gricean implicatures. He introduces a third category of contents, intermediate between what is said and what is implicated, that he labels 'implicatures.' He regards these as contents that are implicit in what is said and that require pragmatic processes of either **completion** or **expansion** to be made explicit. Consider the following, where the content in brackets is supplied from contextually available information:

- (8) Mary swallowed {the bug that flew into her mouth}.
- (9) Mary invited everyone {in her department} to her wedding.

The sentence 'Mary swallowed' is syntactically and semantically complete. (Compare it to 'Mary ate,' which is syntactically but not semantically complete, since 'eat' is a two-place relation. Or to 'Mary devoured,' which is neither syntactically nor semantically complete, since 'to devour' subcategorizes for an obligatory second NP, and 'devour' has two semantic arguments). Bach regards (8) as an example of **conceptual** incompleteness, and hence a pragmatic process of completion must operate, resulting in the derivation of the implicature that Mary swallowed the bug that flew into her mouth.

On the other hand, 'Mary invited everyone to her wedding' expresses a complete proposition. Bach calls this a **minimal proposition**, since the domain of the quantifier is not restricted in any way (beyond the restriction to persons that is encoded by 'one' in 'everyone'). However, this minimal proposition is not the one that the speaker intends to communicate. A pragmatic process of expansion is required, yielding the implicature that Mary invited everyone in her department to her wedding. Bach calls examples such as (9) cases of **sentence nonliterality**. No expression in the sentence is used nonliterally, yet the minimal proposition expressed by the sentence is not what the speaker intends to convey.

Bach denies that there are pragmatically determined aspects of what is said. What is said for Bach is a more minimal notion, which is tied to explicitly encoded semantic content. Bach in effect accepts what Carston (1988) calls the Linguistic Direction Principle. The only contextually determined content that belongs to what is said by the utterance of a sentence is content that corresponds to some element that is

syntactically realized in that sentence. Thus the contextual values of the indexicals in 'She is swallowing now' will be part of what is said by an utterance of this sentence, but the implicit content that specifies what was swallowed (if anything) will not be a part of what is said, since that content corresponds to no element in the sentence. Bach's minimalism requires him to admit that on some occasions what a speaker says does not correspond to a complete proposition. Such is the case in example (8) above. In such cases Bach argues that what is said corresponds to a 'propositional radical,' a gappy object whose missing conceptual elements must be supplied by the context.

Each of the authors discussed above posits a different view of the boundary between semantics and pragmatics. According to Grice, sentence meaning, truth-conditional content, and what is said are all aligned and fall on the side of semantics, whereas implicatures fall on the side of pragmatics. Cohen basically preserves Grice's dichotomy. Cases that may seem to be pragmatic intrusions into truth-conditional content are instead incorporated into semantics. If there is a challenge to Grice it is simply that some phenomena that Grice would label as conversational implicatures are reanalyzed by Cohen as part of semantically encoded content, so that the domain of pragmatics shrinks.

Subsequent views can all in one way or another be seen to challenge Grice's neat dichotomy. Bach remains the most faithful to Grice, since on the whole he preserves the alignment of sentence meaning, truth-conditional content, and what is said on the side of semantics. Truth-conditional content may sometimes come apart from what is said, in those cases in which what is said is conceptually incomplete and hence does not correspond to a complete, truth-evaluable proposition. But when we have truth-conditional content, it is something that is delivered purely by semantics. However, Bach argues that Grice's view of what lies on the side of pragmatics is inadequate. The phenomena of semantic underspecification and sentence nonliterality require us to recognize a category of pragmatic content intermediate between what is said and what is implicated – the category of implicatures.

Levinson's (1995, 2000) views are also quite close to Grice's. He accepts that sentence meaning and what is said line up, and that these are semantic phenomena. However, he allows that there can be pragmatic intrusion into truth-conditional content, so this notion is not a purely semantic notion. On the other hand, Levinson's conception of pragmatics is conservative. He does not challenge the adequacy of the Gricean conception of pragmatics as the domain of conversational implicatures. He does,

however, develop Grice's notion of a generalized conversational implicature (GCI) to a substantial degree. GCIs are said to be default meanings, which belong to a third level of meaning that Levinson calls utterance-level meaning, different from either sentence meaning or speaker meaning. It is only GCIs that are involved in pragmatic intrusion.

Carston's and Recanati's challenges to Grice are more radical. For them, the only purely semantic notion is sentence or expression meaning. What is said (which is equated with truth-conditional content) falls on the side of pragmatics, since there is pragmatic intrusion into truth-conditional content. (Remember, Carston prefers the term 'explicature'.) Carston (2002) suggests that we make a distinction between lexical and truth-conditional semantics. Lexical semantics studies those aspects of meaning that have some sort of syntactic reflex in the language and hence that are a part of the mental lexicon. The lexicon is a store of words in long-term memory. An entry in the mental lexicon is in effect a rule correlating the phonological, syntactic, and semantic information associated with a word. Lexical semantics on this conception is a mentalist enterprise. We can continue to talk of truth-conditional semantics, so long as we realize that the project is very different from the traditional one. Truth-conditions are not assigned directly to the sentences of a language, since sentences by themselves do not have truth-conditions. It is only sentences as used by speakers in particular conversational contexts that have truth-conditions. For a defense of a similar claim, see Stainton (2000).

Moreover, both Carston and Recanati reject Gricean pragmatics as inadequate, since for them pragmatics is not confined to the study of conversational implicatures. Recanati's rejection of Gricean pragmatics may be the most thoroughgoing, since for him, the pragmatic processes involved in the recovery of what is said are not even of the same type as the global pragmatic processes involved in the recovery of implicatures. They are noninferential processes.

Many other voices have been added to this debate. For example, Stainton (1995) argues for the view that semantic underspecification and pragmatic intrusion is rife. He points to the fact that many utterances are of sentence fragments, rather than of complete sentences. Consider:

(10) Top shelf.

Suppose Mary is making herself a sandwich and is rooting around in the kitchen cupboard looking for jam to spread on her toast, and that this is mutually manifest to Mary and her husband Peter. Peter could

utter sentence fragment (10), meaning to convey the proposition that the jam that Mary is looking for is on the top shelf of the cupboard she is searching in. Stainton argues that cases such as these are not to be treated as cases of ellipsis. The missing content in (10) need not correspond to any well-defined syntactic element, as happens in standard cases of syntactic ellipsis, such as the VP-ellipsis in 'Mary donated blood and so did Peter.' (Ellipsis is a very vexed subject. Whether it is something that can be handled in the syntax is not at all clear. See Jackendoff, 1997: 75–78, for a discussion of some problematic cases.) We should accept that language understanding is able to proceed on the basis of fragmentary clues from semantically decoded content. A large burden is placed on the inferential capacities of hearers, who must elaborate these clues on the basis of contextually available information.

Stanley (2002) argues for a diametrically opposed view. According to Stanley, there is much more syntactic and semantic structure than meets the eye, and many of the alleged cases of semantic underdetermination calling for pragmatic enrichment can be re-analyzed as cases where some hidden element in the underlying sentential structure is directing the process of content retrieval. In other words, we preserve the idea of the linguistic direction of content, although the elements doing the directing are often hidden elements (ones that are not phonetically realized, although they are a part of underlying logical form).

Stanley's views have been especially influential in accounting for cases of quantifier domain restriction, such as (9) above or (11) below:

(11) Every child has been vaccinated against polio.

(12) In every country, every child has been vaccinated against polio.

It is possible to use (11) in an appropriate context to convey the proposition that every child **in the United States** has been vaccinated against polio. Stanley's view is that there must be a hidden free variable in (11) whose value is specified in that context as the United States. This variable is present in the underlying logical form of (11) but is not phonetically realized. (Strictly speaking, what is implicit is a free **function** variable, and what must be specified in context is both the function and the values of the arguments of this function. In (11) the function is something like 'resident-in(x).') The evidence that there is a hidden free variable in (11) is that this variable can sometimes be bound by a quantifier. For example, when (11) is embedded in a sentence with a quantifier that has wide scope over the embedded quantifier, as is the case in (12), what is said is that in every country *x*, every child *in x* has been

vaccinated. For more detailed arguments, both pro and con, see Stanley and Szabó (2000), Bach (2000) and Neale (2000). This hidden indexical view has consequences for the quantificational analysis of definite descriptions and has led to some controversy as to the correct analysis of so-called incomplete descriptions. See the chapters in Part I of Reimer & Bezuidenhout (2004).

The Mentalist Picture of the Semantics-Pragmatics Boundary

It was noted above that many philosophers follow Frege in rejecting mentalist semantics. Many linguists on the other hand embrace mentalism, and in fact regard it as the only sensible perspective from which to study language. See Chomsky (2000), Jackendoff (1997, 2002). From the point of view of mentalism, the dispute about the semantics-pragmatics boundary is not one about how to delineate the notions of explicature, implicature, etc. Rather, it is concerned with the question as to how semantic and pragmatic knowledge are represented and organized in the human mind/brain and how this information is combined in the course of on-line production and comprehension of language. Several of the authors discussed above straddle the divide between philosophy and linguistics. Carston, for instance, works in the mentalist tradition. She is concerned to offer a mentalist theory of language **performance**. (So the suggestion made at the outset that semantics is concerned with competence and pragmatics with performance is one that Carston would reject.) On the other hand, although she is interested in articulating a cognitive theory of performance, she has also been an active contributor to the philosophical debate about how to delineate what is said from what is conversationally implied.

Within the mentalist framework, the dominant picture of the semantics-pragmatics interface has been that it is the interface between the language system proper and what Chomsky (1995) calls the conceptual-intentional system. From the comprehension perspective, this interface is where the output from the hearer's language system, namely a representation of the logical form (LF) of the speaker's utterance, is interpreted by processing it in the context of currently active pragmatic information, including information about the speaker's likely communicative intentions. From the production perspective, this interface is where the process of giving expression to a speaker's communicative intentions is initiated. Appropriate lexical-conceptual entries in the speaker's mental lexicon are accessed, thus initiating a process that will ultimately result in the output

of an appropriate phonetic form (PF) at the interface between the language system and the articulatory system.

Jackendoff (1997, 2002) challenges this Chomskyan view while remaining within the mentalist camp. He argues that the language system has a tripartite parallel architecture. There are three independent generative systems or modules, the phonological system, the syntactic system, and the conceptual system. Each contains its own compositional rules and proprietary set of representations, of, respectively, phonological structure (PS), syntactic structure (SS), and conceptual structure (CS). However, it is necessary for these systems to communicate with one another, and they do this via various interface modules, whose job it is to map representations from one system into the representations of another. There is a PS-SS interface, an SS-CS interface (which Jackendoff calls the syntax-semantics interface), and a PS-CS interface. The lexicon is also an interface module, and the interfaces already mentioned are in effect parts of this larger interface system. The lexicon is a long-term memory store whose entries are triples of the three sorts of structures mentioned, namely of PS, SS, and CS. The lexical entry for expression α , $\langle PS_\alpha, SS_\alpha, CS_\alpha \rangle$, is in essence a correspondence rule mapping representations from the three systems into each other. (Lexical entries may be for words, for phrases, such as idioms, or for expressions below the word level, such as agreement markers.)

What Jackendoff calls the syntax-semantics interface (namely the SS-CS interface) is of relevance to the current discussion, since he justifies his claim that CS is the level of semantics, and that it is a level separate from syntax, by appeal to phenomena of the sort that Carston, Recanati, and others appeal to in arguing for pragmatic intrusion into what is said (i.e., into the proposition expressed by an utterance). Jackendoff (1997) argues for what he calls enriched semantic composition. At the level of CS, the compositional principles that form propositions (or thoughts) are sensitive to information that comes from the pragmatic context. But not all this conceptual structure is reflected in the corresponding syntactic structures. Consider the following:

- (13) Peter kept crossing the street.
- (14) Mary finished the book.
- (15) The ham sandwich wants his check.

(13) illustrates the process of aspectual coercion, (14) of co-composition, and (15) of pragmatic transfer. See Pustejovsky (1995) for an account of the first two processes and Nunberg (1979) for an account of the third.

A single act of crossing the street is not a repetitive and (potentially) open-ended action like clapping one's hands or bouncing or spinning a ball. But 'kept' requires such repetitiveness and/or open-endedness. Thus, in the case of (13), 'kept' coerces an interpretation of 'crossing the street' according to which there is either a repeated or an extended action. That is, either we understand Peter to have crossed the street multiple times, perhaps in his effort to lose the detective tailing him. Alternatively, we zoom in on Peter's action of crossing the street and see it as one whose end point is still in Peter's future. Perhaps the street is a very broad one, with a median strip, where Peter pauses briefly before continuing with his crossing.

In the case of (14), finishing is something that can be predicated of an event, but 'the book' refers to an object, not an event. Pustejovsky (1995) argues that the lexical-conceptual entry for 'book' contains information about the typical features of books, namely that they have authors, are read by readers, etc. This conceptual information, presumably along with other contextual information, can be used to arrive at an enriched interpretation of (14) according to which Mary finished reading (or writing, or binding, or illustrating, etc.) the book. In the case of (15), contextual knowledge about restaurants and what goes on in them is used to arrive at an interpretation according to which the ham sandwich orderer wants his check.

In all these cases Jackendoff argues that there is more conceptual (semantic) structure than is represented syntactically. Some will be inclined to argue that there must be covert syntactic structure to match the semantic structure – structure that is there but is not phonetically realized. But Jackendoff argues that this is a mistake. Those who argue for covert structure are in the grip of an assumption that he calls syntactocentrism, namely the view that the only source of compositional structure is syntax. This is an assumption he attributes to Chomsky, since it is built into all the theories of the organization of the language system that Chomsky has proposed, from the Standard Theory of the 1960s, through the extended and revised versions of the Standard Theory in the 1970s, to the Government and Binding (GB) approach of the early 1980s and the minimalist approach of the 1990s. But Jackendoff's account of the tripartite parallel architecture of the language system rejects this assumption.

Moreover, Jackendoff goes further and argues that it is a mistake to talk of **any** semantic structure being directly encoded in the syntax, as Chomsky seems to suggest when he introduces the level of logical form (LF) and talks of it as the level in syntax that directly represents meaning (Chomsky, 1986: 68). It is unnecessary and perhaps even incoherent to talk

in this way. First, it is unnecessary, since the correspondence rules belonging to the syntax-semantics interface (the SS-CS interface) will do the work of correlating syntactic and semantic structures. Note also that the correspondence doesn't have to be perfect. There may be only a partial homology between these two systems. If the communicative system as a whole works in such a way that semantic structure is recoverable from readily available contextual knowledge, then the fact that some of this structure is invisible to the syntactic system is no bad thing.

Second, talk of semantic structure being encoded in syntax may be incoherent if that is allowed to mean that semantic distinctions are directly represented in the syntactic system. The syntactic system is a module whose internal operations are defined over representations in its own proprietary code. So the syntactic system knows about nouns and verbs, case markings, active and passive constructions, WH-movement, etc., not about objects and events, predicate-argument structure, the telic/atelic distinction, thematic roles, etc. Thus it could not represent the sort of pragmatic knowledge needed to interpret examples such as (13)–(15).

As already mentioned, Carston accepts Chomsky's picture, including the assumption of syntactocentrism. She holds that the output from the language system is a representation of LF, which includes those semantic features that are directly syntactically encoded. Earlier we saw her acceptance of the idea that lexical semantics is the study of such encoded aspects of meaning. So, for her, the SS-CS interface would be better called the semantics-pragmatics interface, not the syntax-semantics interface. This makes it seem that her views are very far from those of Jackendoff. Yet Carston's notion of pragmatic enrichment and Jackendoff's notion of enriched composition are very similar.

Jackendoff (2002: 273) does briefly allude to what he might call the semantics-pragmatics interface. It turns out to be an interface level between two sub-levels **within** the conceptual system. It is the level that integrates thoughts that are conveyed by means of language with one's previous knowledge, including knowledge of the communicative context and the speaker's intentions. Such integration may lead one to inferentially derive further thoughts (i.e., Gricean implicatures). In other words, Jackendoff's conception is basically the Gricean conception that is rejected by Carston, since it confines pragmatics to the derivation of implicatures, whereas Carston thinks pragmatic processes are also involved in the enrichment of lexical concepts (encoded meanings) to arrive at ad hoc concepts (contextualized meanings). Of course, Jackendoff can use terminology in the way

he pleases. However, to make it clearer that his views are in fact very close to those of Carston, it might be more appropriate to relabel Jackendoff's SS-CS interface the syntax-pragmatics interface.

This does of course still leave some disagreements unsettled. In particular, it leaves unsettled the issue of syntactocentricism and the debate as to whether there is a specifically linguistic part of semantics, separate from nonlinguistic knowledge, thought, and contextualized meaning. (See Jackendoff, 2002: 281–293, for reasons to deny that there is any such level of semantics.)

See also: Character versus Content; Context Principle; Intention and Semantics; Metaphor: Philosophical Theories; Pragmatic Determinants of What Is Said; Sense and Reference: Philosophical Aspects; Speech Acts; Truth Conditional Semantics and Meaning.

Bibliography

- Bach K (1994). 'Conversational implicature.' *Mind & Language* 9, 124–162.
- Bach K (2000). 'Quantification, qualification, and context: a reply to Stanley and Szabó.' *Mind & Language* 15, 262–283.
- Carston R (1988). 'Implicature, explicature and truth-theoretic semantics.' In Kempson R (ed.) *Mental representations: the interface between language and reality*. Cambridge: Cambridge University Press. 155–181.
- Carston R (2002). *Thoughts and utterances*. Oxford: Blackwell.
- Chomsky N (1986). *Knowledge of language*. New York: Praeger.
- Chomsky N (1995). 'Language and nature.' *Mind* 104, 1–61.
- Chomsky N (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Cohen L J (1971). 'Some remarks on Grice's views about the logical particles of natural language.' In Bar-Hillel Y (ed.) *Pragmatics of natural languages*. Dordrecht: Reidel. 50–68.
- Grice P (1975). 'Logic and conversation.' In Cole P & Morgan J (eds.) *Syntax and semantics* 3. New York: Academic Press. 41–58.
- Grice P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Jackendoff R (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff R (2002). *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Kasher A (1991). 'On the pragmatic modules: a lecture.' *Journal of Pragmatics* 16, 381–397.
- Levinson S (1995). 'Three levels of meaning.' In Palmer F R (ed.) *Grammar and meaning*. Cambridge: Cambridge University Press. 90–115.
- Levinson S (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lewis D (1972). 'General semantics.' In Davidson D & Harman G (eds.) *Semantics for natural language*. Dordrecht: Reidel. 169–218.
- Morris C (1938). *Foundations of the theory of signs*. Chicago: University of Chicago Press.
- Neale S (2000). 'On being explicit: comments on Stanley and Szabó, and on Bach.' *Mind & Language* 15, 284–294.
- Nunberg G (1979). 'The non-uniqueness of semantic solutions: polysemy.' *Linguistics and Philosophy* 3, 143–184.
- Prince E (1988). 'Discourse analysis: a part of the study of linguistic competence.' In Newmeyer F J (ed.) *Linguistics: the Cambridge survey 2: Linguistic theory: extensions and implications*. Cambridge: Cambridge University Press. 164–182.
- Prince E (1997). 'On the functions of left dislocation in English discourse.' In Kamio A (ed.) *Directions in functional linguistics*. Amsterdam: John Benjamins. 117–143.
- Pustejovsky J (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Recanati F (1989). 'The pragmatics of what is said.' *Mind and Language* 4, 295–329.
- Recanati F (2004). *Literal meaning*. Cambridge: Cambridge University Press.
- Reimer M & Bezuidenhout A (eds.) (2004). *Descriptions and beyond*. Oxford: Oxford University Press.
- Sperber D & Wilson D (1986). *Relevance: communication and cognition*. Oxford: Blackwell.
- Stainton R (1995). 'Non-sentential assertions and semantic ellipsis.' *Linguistics and Philosophy* 18, 281–296.
- Stainton R (2000). 'The meaning of "sentences."' *Noûs* 34, 441–454.
- Stanley J (2002). 'Making it articulated.' *Mind and Language* 17, 149–168.
- Stanley J & Szabó Z G (2000). 'On quantifier domain restriction.' *Mind & Language* 15, 219–261.

Sense and Reference: Philosophical Aspects

A Sullivan, Memorial University of Newfoundland,
St John's NL, Canada

© 2006 Elsevier Ltd. All rights reserved.

The Origins and Central Core of the Sense/Reference Distinction

In his earliest works (e.g., Frege [1879]), Frege espoused a referentialist approach to language, according to which the meaning of an expression is whatever it is to which the expression refers. The referentialist picture is simple and compelling: singular terms such as 'Nancy' mean objects, general terms such as 'doctor' mean sets or properties, and the meaning of a sentence such as 'Nancy is a doctor' is the fact or state of affairs that the sentence is about.

Informative identity statements – i.e., true but non-trivial statements of the form ' $a = b$ ' – shattered this picture for Frege. His classic example is

- (1) Hesperus is Phosphorus.

According to the lore, the Evening Star (the brightest heavenly body visible in the evening at certain times of the year) was baptized 'Hesperus,' whereas the Morning Star (the brightest body in the dawn sky at certain times) was baptized 'Phosphorus.' With the growth of astronomical knowledge, it turns out that Hesperus *is* Phosphorus, and what is more it is not a star at all but the planet Venus. The problem is to account for the difference between the meanings of (1) and

- (2) Hesperus is Hesperus.

As (1) and (2) are constructed from co-referential parts, the referentialist approach seems to be committed to the view that they say exactly the same thing. However, according to Frege, there are substantive differences here: " $a = a$ ' and ' $a = b$ ' are obviously statements of differing cognitive value; ' $a = a$ ' holds *a priori* ... while statements of the form ' $a = b$ ' often contain very valuable extensions of our knowledge and cannot always be established *a priori*" (1892: 175). Because referentialism cannot account for this evident fact about meaning, it is untenable, and a new view is wanted.

Frege's problem, then, is this: insofar as the function of a term is just to single out its referent, there is no way to distinguish the meanings of co-referential terms, and so no way to distinguish the propositions expressed by sentences that differ only in the interchange of co-referential terms. Frege's solution is that the semantic contribution of a term encompasses

more than just its referent. 'Sense' is the name for the dimension along which the semantic contributions of co-referential terms can differ. Every term is associated with a sense, and the sense specifies the condition for being the term's referent. Co-referential expressions can have distinct senses, and expressions need not have an actual concrete referent to have a sense.

The distinction between sense and reference is clearly illustrated in the case of a definite description, such as 'the tallest woman in Mongolia' or 'the architect who designed this building.' The sense is the identifying condition expressed by the term. It is that which must be grasped in order to understand the expression, and it is accessible to any competent speaker, regardless of whether they have been to Mongolia or seen this building. The referent is the individual that satisfies the condition that the sense specifies. Frege (1892) characterizes the sense as the mode of presentation of the referent. There are multiple ways of describing and conveying information about one and the same referent; and to each of these ways corresponds a distinct sense.

Frege's distinction generalizes well beyond cases like (1) and (2) – it is not limited to identity statements, or to singular terms. For example, (3) and (4) might differ in meaning, and (5) and (6) in truth-value, despite differing only in the interchange of co-referential terms:

- (3) Hesperus is visible in the evening.
- (4) Phosphorus is visible in the evening.
- (5) The ancient Babylonians believed that Hesperus is visible in the evening.
- (6) The ancient Babylonians believed that Phosphorus is visible in the evening.

Some related points are illustrated by sentences such as:

- (7) Pegasus does not exist.

Such negative existential statements with non-denoting names in the subject-position pose problems for referentialism, as (7) is clearly meaningful despite the fact that its subject-expression does not refer to any actual thing. Furthermore, the same points apply to general terms. For example, to be a renate, one must have kidneys, whereas to be a cordate, one must have a heart. All and only renates are cordates, so the terms are co-referential; nonetheless, they clearly differ in meaning, and so are subject to the phenomena illustrated by (1)–(6). There are also non-denoting

general terms – the term ‘phlogiston’, for instance, was introduced to name a substance given off in the process of fire. With the growth of knowledge, it turns out that there is no such thing, and so (8) poses problems for referentialism similar to those posed by (7):

(8) Phlogiston does not exist.

Frege concludes that, in general, the semantic contribution of an expression is not simply its referent. Instead, every significant linguistic expression has associated with it a sense that determines its referent. So, Frege’s distinction amounts to a fully general approach to language that posits an intermediate layer of sense between words and referents.

More on Frege’s Distinction

Frege’s approach to the word-referent relation is a classic instance of what is called ‘mediated’ or ‘indirect’ reference – reference is *indirect* because there is a sense *mediating* the link between a term and its referent. The characteristic virtue of the indirect reference approach is that it affords a clear semantic distinction between co-referential expressions.

The distinction between sense and reference is a distinction between what we say and what we say it about. The sense of an expression is its contribution to the content of the thought or proposition expressed by uses of sentences in which it figures; the referent of an expression is its contribution to the truth-value of this thought or proposition. To illustrate, when enjoying a work of fiction, we are only engaged at the level of sense – “The thought [expressed by a sentence in the *Odyssey*] remains the same whether ‘Odysseus’ has a reference or not” (1892: 180). However, if we are interested in whether a thought is true, then we become engaged at the level of reference. “The question of truth would cause us to abandon aesthetic delight for an attitude of scientific investigation. ... It is the striving for truth that drives us always to advance from the sense to the reference” (1892: 180). At the limiting case of a complete sentence, Frege takes its sense to be a thought or proposition, and its referent to be a truth-value (see Frege [1891, 1892] for further explanation).

Frege’s individuates senses in terms of cognitive significance. Two statements differ in cognitive significance if it is possible for a competent speaker to take contrasting attitudes toward them (i.e., to believe that one is true while disbelieving, or withholding judgment concerning, the second). According to Frege, pairs such as (1)–(2) and (3)–(4) differ in cognitive significance. If two statements that differ only in

the interchange of co-referential terms differ in cognitive significance, this proves that the terms differ in sense.

Consider again (3)–(6). Even though (3) and (4) differ in cognitive significance, they are nonetheless truth-conditionally equivalent, because they are built from co-referential parts. How then can (5) and (6) differ in truth-value, if truth-conditions are solely a matter of referents, not senses? Here Frege (1892) argues that in certain contexts (such as inside the scope of ‘A believes that ...’) the referent of an expression is its customary sense. (5) and (6) assert relations between agents and thoughts, not between agents and truth-values. They attribute distinct thoughts to the Babylonians, and thus can differ in truth-value.

Sense is that which links belief, meaning, and rational action. One might, for instance, believe that Hesperus is populated by an advanced and benevolent race and desire above all else to visit them, but still turn down a free trip to Phosphorus – and not be in the least bit irrational – as long as one does not believe that Hesperus is Phosphorus. Thus, senses are integral to the task of explaining and predicting rational action. Relatedly, merely identifying the referents of the expressions uttered is not sufficient for understanding what a speaker has said, on a Fregean view. (For instance, if one says ‘Hesperus is Phosphorus,’ to report or translate this utterance as saying that Hesperus is Hesperus is clearly unsatisfactory, even though it preserves truth-conditions.) To understand what has been expressed, one also needs to bear in mind the speaker’s perspective on those referents.

Subsequent History, and Criticisms

Frege’s distinction between sense and reference has had enormous impact on the subsequent development of the philosophy of language. Starting with Russell (1905), and continuing through seminal work by Davidson (1967), Kripke (1972), Putnam (1975), and Kaplan (1977), theorists begin their constructive projects by situating their theories in relation to Frege’s. Furthermore, the influence of Frege’s distinction is far from exclusively negative. There is a rich tradition, running through such influential theorists as Carnap (1947), Church (1951), and Montague (1974), which follows Frege’s indirect reference approach in assigning two such semantic entities to every significant linguistic expression. Down this path, Frege’s distinction had a direct influence on the development of more or less every single branch of intensional logic and semantics.

More generally, Frege's distinction is now pervasive throughout philosophy, a part of the canon. Whether the topic be ontology, minds, or morals, in any case in which it is important to bear in mind subtle distinctions between conceptual conditions and that which satisfies those conditions, one is likely to encounter the sense/reference distinction. (This is clearly illustrated in recent philosophy of mind, in which finessing the distinction between the sense and the reference of psychological [qualitative, phenomenal] terms is – interestingly, and tellingly – a core part of the arguments of both dualists [cf. Chalmers (1996)] and the physicalists who oppose them [cf. Papineau (2002)].)

There are, however, many critical questions raised by Frege's distinction between sense and reference. For one thing, positing senses gives rise to difficult metaphysical questions, such as: 'What, exactly, are senses?' Frege (1892, 1918) gives arguments about what senses are not – because co-referential expressions can differ in sense, senses are not of the physical realm; since distinct mental states can have exactly the same sense, senses are not private mental entities. However, as for what senses are, exactly, Frege (1918) just offers a vague metaphor: senses exist in some third realm, distinct from the physical and mental realms. Understandably, many find this 'third realm' talk to be beyond belief. At best, it is in need of significant clarification.

Some of the most influential objections to Frege's distinction were raised in the 1970s – led by Donnellan (1970), Kripke (1972), Putnam (1975), and Kaplan (1977). One way to put their basic thrust is this: two central constraints on senses are that they be answerable to concerns of cognitive significance and that they be objective; however, there is something subjective, or speaker-relative, about cognitive significance. (For example, two competent speakers could disagree as to whether it is significant or trivial to be told that Homer is the author of the *Odyssey*; and it is not evident that there are any objective grounds that would conclusively resolve such a dispute.) Thus, these two constraints pull in opposite directions, and it is not clear that any one thing can satisfy them both. More generally, the aforementioned articles give reasons to be skeptical as to whether expressions have one determinate objective sense, and, even if so, reasons to be skeptical as to whether this sense determines reference.

Led by Dummett (1981), Evans (1982), and McDowell (1986), neo-Fregeans have sought to defend refined interpretations of Frege's doctrines from some of these objections. (One of the core ideas is that many criticisms of Frege's views can be met if senses are conceived as object-dependent – and so

'rigid,' in Kripke's [1972] terminology.) In the other direction, original criticisms of Frege's approach continue to surface. (For an important example, see Schiffer's [1992] argument that nothing can satisfy the demands that Frege places on the notion of 'sense'.) In any case, it is evident that Frege's problem does pose a challenge to many views of meaning and reference, and that there is some aspect of meaning, distinct from reference, that is essential to understanding rational thought and intentional action. Exactly what Frege's problem shows, or exactly how to characterize the latter aspect of meaning, though, are complex matters that remain controversial.

See also: Direct Reference; Empty Names; Proper Names: Philosophical Aspects.

Bibliography

- Carnap R (1947). *Meaning and necessity*. Chicago: University of Chicago Press.
- Chalmers D (1996). *The conscious mind*. Oxford: Oxford University Press.
- Church A (1951). 'Intensional semantics.' In Martinich A (ed.) *The philosophy of language*. Oxford: Oxford University Press, 1985. 40–47.
- Davidson D (1967). 'Truth and meaning.' *Synthese* 17, 304–323.
- Donnellan K (1970). 'Proper names and identifying descriptions.' *Synthese* 21, 256–280.
- Dummett M (1981). *The interpretation of Frege's philosophy*. Cambridge, MA: Harvard University Press.
- Evans G (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Frege G (1879). 'Begriffsschrift.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 93–117.
- Frege G (1891). 'Function and concept.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 143–161.
- Frege G (1892). 'On sense and reference.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 175–192.
- Frege G (1918). 'The thought: a logical inquiry.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 201–218.
- Kaplan D (1977). 'Demonstratives.' In Almog J, Perry J & Wettstein H (eds.) *Themes from Kaplan*. Oxford: Oxford University Press, 1989. 481–564.
- Kripke S (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- McDowell J (1986). 'Singular thought and the extent of inner space.' In MacDowell J & Pettit P (eds.) *Subject, thought, and context*. Oxford: Oxford University Press. 137–168.
- Montague R (1974). *Formal philosophy*. New Haven: Yale University Press.

- Papineau D (2002). *Thinking about consciousness*. Oxford: Clarendon.
- Putnam H (1975). 'The meaning of "meaning."' In Gunderson K (ed.) *Mind, language, and reality*. Cambridge: Cambridge University Press. 131–193.

- Russell B (1905). 'On denoting.' In Sullivan A (ed.) *Logicism and the philosophy of language*. Peterborough: Broadview, 2003. 235–247.
- Schiffer S (1992). 'Belief ascription.' *Journal of Philosophy* 89, 499–521.

Situation Semantics

V Akman, Bilkent University, Ankara, Turkey

© 2006 Elsevier Ltd. All rights reserved.

Situation theory starts with a fundamental observation: reality consists of situations. A situation is a rich object consisting of individuals enjoying various properties and standing in a variety of relations. It is, in a sense, a 'small' world. We always find ourselves in situations. Right now, you, the reader, are in a reading situation. You are, I hope, satisfied with your being in this situation (notice that this is an attitude you have toward this situation). Some months ago, I, the author of this article, was in a writing situation (distributed over time and place).

Situations describe parts of the real world. Information flow is made possible by a network of abstract links between high-order uniformities, that is situation types. One of the distinguishing characteristics of situation theory *vis-à-vis* the traditional account is that information content is invariably context-dependent (Akman and Surav, 1997).

Situation semantics is applied situation theory. We are engaged in situation semantics if we are using situation-theoretic ideas – mathematical theories of information content – to study meaning in natural language. In fact, the two areas are not clearly separable, as the still-popular acronym STASS (situation theory and situation semantics) neatly shows.

Unlike the older and widely known approaches to natural language meaning (e.g., Montague grammar), there is certain natural feel to situation semantics. This makes it enticing for a newcomer to the realm of semantics. Situation semantics does not impose human-made assumptions in our conceptual scheme. It may be burdensome for someone to embrace, say, Montagovian intensions, but situations have a certain conceptual clarity and naturalness that make them believable. You may have heard that the classical model theory is a 'model' theory in the sense that it depicts how a logical theory should be like. Likewise, situation semantics is a fine exemplar of what a naturalized theory of semantics should be like.

Classical approaches to semantics underestimate the role played by context; they ignore factors such as intentions and circumstances of the individuals involved in the communicative process. (Or, rather, they place them in the pragmatics basket.) But, linguistic devices such as indexicals, demonstratives, and tenses rely heavily on context for interpretation and are fundamental to the way language carries information. Context-dependence is essential to situation semantics. (The insistence of situation semantics on contextual interpretation makes it compatible with speech act theory and discourse pragmatics.) A sentence can be used over and over again in different situations to say different things (called the efficiency of language). Its interpretation (i.e., the class of situations described by the sentence) is therefore subordinate to the situation in which the sentence is used. This context-providing situation (discourse situation) is the speech situation, including the speaker, the addressee, the time and place of the utterance, and the expression uttered. Because speakers are always in different situations, having different causal connections to the world and different information, the information conveyed by an utterance will be relative to its speaker and hearer (called the perspectival relativity of language).

Context supports not only facts about speakers, addressees, and so on, but also facts about the relations of discourse participants to other contextually relevant situations such as resource situations. Imagine two card games that are going on, one across town from the other. Suppose Alice is playing rummy with Bob and Carol is playing rummy with David. Elwood, watching the former card game, mistakes Alice for Carol, and mutters, *Carol has the ace of clubs*. According to the classical theory, if Carol indeed has the ace of clubs (A♣), his claim would be true since *Carol* and *the ace of clubs* are used to pick, among all the things in the world, the unique objects satisfying the properties of being someone named Carol and being an A♣, respectively. In contrast, situation semantics identifies these objects with respect to some limited situation – the resource situation exploited by Elwood. The claim would then be wrong even if Carol had the A♣ in the other card game.

In traditional semantics, statements that are true in the same models convey the same information. Situation semantics takes the view that logically equivalent sentences need not have the same subject matter because they need not describe situations involving the same objects and properties. The notion of partiality leads to a more fine-grained notion of information content and a stronger notion of logical consequence that does not lose track of the subject matter.

Ambiguity is another aspect of the efficiency of language. Natural language expressions may have more than one meaning. There are factors such as intonation, gesture, the place of an utterance, and so on that may play key roles in the interpretation of an utterance. Instead of downgrading ambiguity as an impurity of natural languages, situation semantics tries to build it into a full-fledged theory of linguistic meaning.

Intelligent agents generally make their way in the world by being able to pick up certain information from a situation, process it, and react accordingly. Being in a situation, such an agent has information about the situations he or she sees, hears about, believes in, and so on. Thus, upon hearing Bob's utterance *a wolf is running toward you*, Alice would have the information that her friend is the speaker and that he is addressing her with *you*. Moreover, by relying on the situation the utterance described, she would know that there is a wolf fast approaching her. She would then form a thought about this – an abstract object having the property of being a running wolf – and, on seeing the wolf around, her thought would start to correspond with facts. Normally, the realization of some type of situation causes an agent to acquire more information about that situation and to act accordingly. Alice would run away, having in her possession the acquired knowledge that wolves are hazardous. She activates this knowledge from the situation she finds herself in via a constraint – the link between wolves and their fame as life-threatening creatures. The role of constraints in information flow is best illustrated with an example. The statement *smoke means fire* expresses the lawlike relation that links situations in which there is smoke to situations in which there is a fire. If s is the type of smoky situations and f is the type of fire situations, then, by being attuned to the constraint $s \Rightarrow f$, an agent can pick up the information that there is a fire in a particular situation by observing that there is smoke.

Meaningful expressions are used to convey information not only about the external world but also about our minds (called the mental significance of language). Returning to an earlier example, consider

the sentence *a wolf is running toward you* uttered by Bob. It can give Alice information about two different situations. The first one is the situation that she is located in. The second one is Bob's mental (belief) situation. If Alice is certain that he is hallucinating, then she learns the second situation, not the first. Situation semantics differs from other approaches in that in attitude reports we do not describe our mind directly (by referring to states of mind, ideas, senses, thoughts, and whatnot) but indirectly (by referring to situations that are external).

According to situation semantics, the meanings of expressions reside in systematic relations between different types of situations. They can be identified with relations on discourse situations d , (speaker) connections c , the utterance situation u itself, and the described situation e . Some public facts about u – such as its speaker and time of utterance – are determined by the discourse situations. The ties of the mental states of the speaker and the hearer with the world constitute c .

A discourse situation d involves the expression uttered, its speaker, the spatiotemporal location of the utterance, and the addressee. Each of these defines a linguistic role (the role of the speaker, the role of the addressee, etc.). The utterance situation u constrains the world in a certain way, depending on how the roles for discourse situations, connections, and described situation are to be filled. For instance, an utterance *I am crying* defines a meaning relation:

$$d, c[[I \text{ am crying}]]e$$

Given a discourse situation d , connections c , and a described situation e , this holds just in case there is a location L and a speaker s such that s is speaking at L , and, in e , s is crying at L .

In interpreting the utterance of an expression f in context, there is a flow of information, partly from the linguistic form encoded in f and partly from contextual factors provided by the utterance situation u . These are combined to form a set of constraints on the described situation e . This situation is not uniquely determined; there may be several situations that satisfy the constraints. The meaning of an utterance of f and hence its interpretation are influenced by other factors such as stress, modality, and intonation. However, the situation in which f is uttered and the situation e described by this utterance seem to play the most influential roles.

Guide to Literature

Ground-breaking work on STASS is due to the late Jon Barwise, well-known mathematical logician,

and John Perry, prominent philosopher of language and mind. Barwise and Perry were the founders of Stanford University's Center for the Study of Language and Information (CSLI), which became almost synonymous with STASS research. In the beginning, the development of situation theory was hampered by a lack of appropriate tools. Later, the theory assembled its foundations based on innovations coming from set theory (Barwise and Etchemendy, 1987; Aczel, 1988). Barwise and Seligman (1997) further advanced the theory by introducing the concept of an information channel, which preserves information as it is transmitted through a system. Their work is in the spirit of Dretske's (1981) landmark work on information flow.

It is impossible to do justice to the profundity of STASS in a brief summary of this kind. The reader is referred to two seminal books, Barwise and Perry (1983) and Devlin (1991), for a thorough understanding. Although somewhat dated, the former is densely packed with excellent semantic common sense. The latter volume proposes a streamlined vocabulary and pays close attention to the foundations; it is the only modern introduction to STASS. (However, it does not render the Barwise and Perry volume obsolete; each book has its own merits.) Seligman and Moss (1997) is a beneficial survey that is mathematically demanding; it also has an excellent bibliography.

Various versions of situation theory have been applied to a number of linguistic issues (mainly) in English. The ideas emerging from research in situation semantics have also been coalesced with well-developed linguistic theories and have led to rigorous formalisms (Fenstad *et al.*, 1987). On the other hand, situation semantics has been compared to another influential approach to the theory of meaning, discourse representation theory (DRT).

Indexicals, demonstratives, referential uses of definite descriptions, deictic uses of pronouns, tense markers, and names all have technical treatments in situation semantics. Gawron and Peters (1990) focused on the semantics of pronominal anaphora and quantification. They argued that the ambiguities of sentences with pronouns can be resolved with an approach that represents anaphoric relations syntactically. They use a relational framework that considers anaphoric relations as relations between utterances in context. Cooper (1991, 1996) offered painstaking studies of linguistic problems to which situation semantics has been applied with some success. Tin and Akman (1994, 1996) showed how situation theory can be given a computational twist. They implemented a prefatory prototype (named BABY-SIT for obvious reasons) to program some practical

problems, including anaphora resolution. Devlin and Rosenberg (1996) explored applications of situation theory to the study of language use in everyday communication to improve human–computer interaction.

There used to be a specialized series of conferences devoted to recent developments in STASS. The first three volumes of proceedings were published as Cooper *et al.* (1990), Barwise *et al.* (1991), and Aczel *et al.* (1993). Nowadays, it is possible to find situation–theoretic papers dispersed in numerous conferences on logic, language, and information.

Finally, Devlin (2004) provides – despite the specific sounding title – a general appraisal of what STASS is all about; it may be consulted to get a better grasp of the historical developments that shaped STASS. Many of the Barwise papers reviewed by Devlin can be found in Barwise (1989), a fertile collection for technically oriented readers.

See also: Anaphora: Philosophical Aspects; Conditionals; Context and Common Ground; Discourse Representation Theory; Expression Meaning versus Utterance/Speaker Meaning; Formal Semantics; Indexicality: Philosophical Aspects; Logic and Language: Philosophical Aspects; Logical Consequence; Meaning: Overview of Philosophical Theories; Metaphysics, Substitution *Salva Veritate* and the Slingshot Argument; Montague Semantics; Possible Worlds: Philosophical Theories; Pragmatic Determinants of What Is Said; Quantifiers: Semantics; Relevance Theory; Representation in Language and Mind; Semantics–Pragmatics Boundary; Sense and Reference: Philosophical Aspects; Speech Acts; Thought and Language: Philosophical Aspects.

Bibliography

- Aczel P (1988). *Non-well-founded sets*. Stanford, CA: CSLI Publications.
- Aczel P, Israel D, Katagiri Y & Peters S (eds.) (1993). *Situation theory and its applications* (vol. 3). Stanford, CA: CSLI Publications.
- Akman V & Surav M (1997). 'The use of situation theory in context modeling.' *Computational Intelligence* 13, 427–438.
- Barwise J (1989). *The situation in logic*. Stanford, CA: CSLI Publications.
- Barwise J & Etchemendy J (1987). *The liar: an essay on truth and circularity*. New York: Oxford University Press.
- Barwise J & Perry J (1983). *Situations and attitudes*. Cambridge, MA: MIT Press.
- Barwise J & Seligman J (1997). *Information flow: the logic of distributed systems*. Cambridge, UK: Cambridge University Press.
- Barwise J, Gawron J M, Plotkin G & Tutiya S (eds.) (1991). *Situation theory and its applications* (vol. 2). Stanford, CA: CSLI Publications.

- Cooper R (1991). 'Three lectures on situation theoretic grammar.' In Filgueiras M, Damas L, Moreira N & Tomás A P (eds.) *Natural language processing*. Berlin: Springer-Verlag. 102–140.
- Cooper R (1996). 'The role of situations in generalized quantifiers.' In Lappin S (ed.) *The handbook of contemporary semantic theory*. Cambridge, MA: Blackwell Publishers. 65–86.
- Cooper R, Mukai K & Perry J (eds.) (1990). *Situation theory and its applications* (vol. 1). Stanford, CA: CSLI Publications.
- Devlin K (1991). *Logic and information*. New York: Cambridge University Press.
- Devlin K (2004). 'Jon Barwise's papers on natural language semantics.' *Bulletin of Symbolic Logic* 10, 54–85.
- Devlin K & Rosenberg D (1996). *Language at work: analyzing communication breakdown in the workplace to inform systems design*. Stanford, CA: CSLI Publications.
- Dretske F (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Fenstad J E, Halvorsen P-K, Langholm T & van Benthem J (1987). *Situations, language, and logic*. Dordrecht, Netherlands: Reidel.
- Gawron J M & Peters S (1990). *Anaphora and quantification in situation semantics*. Stanford, CA: CSLI Publications.
- Seligman J & Moss L S (1997). 'Situation theory.' In van Benthem J & ter Meulen A (eds.) *Handbook of logic and language*. Amsterdam: Elsevier Science. 239–309.
- Tin E & Akman V (1994). 'BABY-SIT – towards a situation-theoretic computational environment.' In Martín-Vide C (ed.) *Current issues in mathematical linguistics*. Amsterdam: North-Holland. 299–308.
- Tin E & Akman V (1996). 'Information-oriented computation with BABY-SIT.' In Seligman J & Westerståhl D (eds.) *Logic, language, and computation*, vol. 1. Stanford, CA: CSLI Publications. 19–34.

Social Construction and Language

D von Daniels, University of Leipzig

© 2006 Elsevier Ltd. All rights reserved.

Nonanalytic Approaches

According to Georg Wilhelm Friedrich Hegel the topic of "language and social construction" encompasses everything there is. I note this all embracing approach to contrast it with how these topics are discussed today. This article will focus on one view concerning the relation between language and social construction, namely John Searle's approach, which at the same time will serve to introduce into the fundamental problems. It will be necessary, however, to show first how this specific way of situating and addressing the problem evolved.

Subsequent to Hegel one can distinguish roughly four lines of thought: (1) the social theorists, ranging from Karl Marx through Emile Durkheim to Max Weber and modern sociology, (2) hermeneutical philosophers, from Wilhelm Dilthey through Edmund Husserl and Maurice Merleau-Ponty to modern post-structuralism, (3) the analytic philosophers starting with Gottlob Frege, Bertrand Russell, and Ludwig Wittgenstein up to modern analytical philosophy. Specific problems of social constructions are (4) also discussed in various disciplines such as jurisprudence, economics, and political science, sometimes accompanied by philosophical inclinations. However, none of these traditions is totally exclusive of the others. Even though the topic 'language and social

construction' features in all such discussions, this article will only consider the analytical tradition.

Analytic Approaches

Early Analytic Approaches and Social Construction

The second generation of Analytic philosophers were in the beginning firmly rooted in either logical positivism or ordinary language philosophy. Their original objectives were to reveal the logical structure of language or to solve philosophical puzzles by analyzing the ordinary usage of words. Both schools of thought shared an empiricist and scientific attitude. For this early reason, analytical philosophers showed little interest in social and political reality or history. A few philosophers within the analytic tradition did pay attention to institutions, however, most notably Peter Winch (1958) and David Lewis (1969). Winch argued on the basis of Wittgenstein's *Philosophical Investigations* that explaining human action requires an understanding of the forms of life established in the society wherein the action is carried out; hence objective, culture independent explanations like those found in the natural sciences are unavailable in the social sciences. David Lewis, on the other hand, gave a game theoretic account of conventions as a solution to recurrent coordination problems. While Winch tried to justify the traditional distinction between the approaches in the natural sciences and the social sciences, Lewis examined ways of reconstructing classical arguments of social philosophy using modern

means. Recently John Searle has developed an ontology of social facts and, in particular, of social institutions based on his speech act theory and philosophy of mind (Searle, 1995). This development marks the beginning of a renewed interest in social facts within analytic philosophy.

John Searle's Approach

Social Construction John Searle's first major work was on speech acts, where he first pointed out the distinction between brute facts and institutional facts, and highlighted the particular role of constitutive rules in the creation of institutions (Searle, 1969: 50–53). He subsequently based the foundations of speech act theory on his philosophy of mind (Searle, 1983). Searle then returned to institutional facts to elaborate his point more systematically. He uses three concepts as building blocks for his theory: the assignment of functions, collective intentionality, and constitutive rules of the form “X counts as Y in context C” (Searle, 1995: 13). Taking the “scientific worldview” for granted, Searle starts with the puzzle of how it is possible that there are objective social facts, such as “George Washington was the first president of the U.S.,” within a world composed entirely of minute physical particles. Searle's basic proposal is that we assign functions both to naturally occurring objects (e.g., using a stone as a paperweight) and to those created for the purpose of performing an assigned function (e.g., using a hammer) (Searle, 1995: 13–23). Searle argues that such functions are never discovered in nature, like the natural properties of weight, color, etc., and cannot be completely reduced to causal explanations. Instead they are always imposed on things by intentional actors. Functions thus are always observer relative. This distinction is crucial because it allows us to mark the difference between features of reality that are observer independent (e.g., gravity, force) and usually described by the natural sciences and those that are observer dependent and usually dealt with by the social sciences. The latter features are always partly constituted by an ontologically subjective set of attributes (intentions); nevertheless they can in the case of institutions still be epistemically objective.

Searle also uses the imposition of functions to explain the meaning of words and signs. According to Searle, meaning arises whenever we assign functions to things as to symbolize, represent, or stand for something. Hence a red traffic light signals ‘Stop!’ because we have assigned this function to red lights in boxes at the roadside. Searle's theory of meaning is taken for granted. The relevant point here concerns only the relation between the natural world and the social world of animals and humans. While Searle

attempts to adhere to the program of realism he nevertheless makes frequent use of terms like intentions, functions, etc. According to Searle they are compatible with the scientific worldview of realism since he claims intentionality, the prerequisite for the imposition of functions, is a “higher level feature of the brain.”

Searle uses the same scheme of explanation to introduce the second building block of his theory, collective intentionality (Searle, 1995: 23–29). He sees collective intentionality or ‘we-intentionality’ as a “biologically primitive phenomenon” that cannot be reduced or eliminated in favor of something else. Searle, in particular, stipulates that it cannot be reduced to “individual intentions plus something else.” Whenever people engage in cooperative social action, Searle argues, there has to be “collective intentionality at a higher level.” This concept holds true for a violinist playing in an orchestra and even for certain hostile activities like litigants in a legal proceeding or two faculty members trading insults at a cocktail party. Searle holds that all these are social facts that involve the ‘we intentions’ of the parties.

Finally, constitutive rules of the form “X counts as Y in the context C” are employed to explain the difference between brute facts and institutional facts (Searle, 1995: 27–29). According to Searle, brute facts, such as that gold is a chemical element, exist independently of social institutions. Only the sentence “gold is an element” depends on the institutions of language and science. Institutional facts, on the other hand, such as the scoring rules of football, depend for their very existence on conventional agreements and on the acceptance of such conventions by the group taking part in the game; they are not pre-given. To explain institutional facts Searle further distinguishes between regulative and constitutive rules. Regulative rules govern preexisting activities, like driving on the right side of the road. The activity of driving (on either side of the road) existed before it was made obligatory to drive on the right. Constitutive rules, on the other hand, create the very possibility of an activity. Without the rules of football there would be no such game, and kicking a ball between two posts would not count as scoring. Searle concedes that the distinction between regulative rules and constitutive rules cannot always be strictly drawn as many constitutive rules rest on preexisting activities that might be governed by regulative rules. In any case, it is a distinction within the realm of social facts. All institutions are then built by using a system of constitutive rules of the form “X counts as Y in the context C” (Searle, 1995: 43–51). Thus, shooting a ball between posts (X) counts as scoring (Y) in a football game (C). The X term stands for some object

or activity, to which (via collective intentionality) a certain status function has been assigned.

The most obvious example of social facts, explainable along the lines of Searle's analysis, is an institution such as money. Status functions are assigned on certain pieces of paper. Such statements require an involvement of collective intentionality; hence the paper functions as money only as long as we consider it to be money. It is moreover dependent on a specific context. The formula is useful for explaining standard cases, though it allows as well for the reflection of unusual cases. A standard case might be: 'Bills issued by the Bureau of Engraving and Printing (X) count as money (Y) in the United States (C).' The existence of a 'cigarette currency' in post-war Germany might be an unusual case: 'American cigarettes (X) counted as money (Y) in post-war Germany (C).' With these basic distinctions at hand it is possible to give a 'hierarchical taxonomy of (certain types of) facts' starting from brute facts and going up to institutional facts (Searle, 1995: 37–43).

The Role of Language Searle's approach makes obvious why social construction is a linguistic activity. Collective intentionality alone does not require language or any symbolic means as it may be presupposed that even a pack of wolves hunting together are engaged in a collective enterprise. Indeed, Searle assumes that animals have intentionality. The difference comes only with the imposition of the status functions.

In moving from the X to the Y term the matter of the X term does not change, it only acquires a new meaning. It is used to symbolize something that exists only by virtue of a collective agreement. This theory requires some elucidation. Symbols are often (but not always) used to refer to objects that exist independently of the symbols and even independently of language. The symbol 'cat' refers to cats; and cats exist independently of any symbol referring to them. The imposition of status functions on something, on the other hand, creates a new category of objects or actions that have not existed prior to symbolization and collective agreement. The difference is made obvious by comparing social behavior of animals and humans. Certain animals might live in lifelong pair-bonding but they will nevertheless not be married, as they do not have the institution of marriage. That humans are married cannot be read off their behavior either but has to be declared in a certain situation. The declaration of marriage is an indispensable linguistic activity as only by virtue of the symbolic act are certain rights and duties granted. Institutions hence create reasons for action that depend on

forms of symbolization. The initial act is usually thereafter represented by some symbolic means, Searle calls them status indicators, as e.g., wedding rings or a marriage certificate.

A further sign evincing their linguistic nature is that status functions are intensional and cannot be replaced *salva veritate* by coextensive expressions. Hence, even if it is true that 'marriage ruins the sex life,' it does not follow that 'exchanging rings counts as "ruining sex life" in the context of a church ceremony.'

It should be kept in mind that the formula is meant not only to explain the creation of institutions out of brute facts but as a general explanation. Therefore, institutional facts are permissible at the place of the X term so that social facts can be reiterated. Indeed, Searle notes that language itself is an institutional fact that can be analyzed using the formula.

However, this feature reveals a fundamental problem. One of Searle's main objectives is to show how institutional facts emerge out of natural facts through language. If, however, language is an institutional fact in itself, the formula merely highlights a structure within the realm of institutional facts but cannot ultimately explain the emergence of institutional facts. Searle addresses the problem merely by stating that language is a special kind of institutional fact, namely one that is "designed to be a self-identifying category of institutional fact" (Searle, 1995: 73).

Critical Assessment

Searle labels his approach "philosophy of society," analogous to classic fields like philosophy of language or philosophy of mind. Three kinds of criticisms in reaction to Searle's proposal have emerged. Those who agree in principle with Searle's line of thought dispute specifically the concepts of collective action and collective intentionality (Gilbert, 1989; Bratman, 1999; Tuomela, 2002). There are, on the other hand, two kinds of criticisms questioning Searle's approach in principle. Some philosophers hold that Searle misses the fundamental role ordinary language plays in conceiving reality at all, since Searle presupposes implicitly that the world as described by the natural sciences can be attained independently and prior to the everyday social world, hence independently of ordinary language. Social scientists, on the other hand, criticize Searle for failing to account for social facts properly, since social facts pertain to the explanation of actions, a feature Searle omits in his philosophy of society. The different kinds of criticism do not exactly correspond with the lines of thought mentioned in the beginning; however they show that the philosophical discussion is still divided over fundamental questions.

See also: Analytic Philosophy; Philosophy of Science and Linguistics.

Bibliography

- Anscombe G E M (1958). 'On brute facts.' *Analysis* 18, 69–72.
- Bratman M (1999). *Faces of intention: selected essays on intention and agency*. Cambridge: Cambridge University Press.
- Gilbert M (1989). *On Social Facts*. London: Routledge.
- Lewis D K (1969). *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.

- Searle J R (1969). *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Searle J R (1983). *Intentionality. An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Searle J R (1995). *The construction of social reality*. New York: The Free Press.
- Tuomela R (1994). *The importance of us: a philosophical study of basic social notions*. Stanford, CA: Stanford University Press.
- Tuomela R (2002). *The philosophy of social practices: a collective acceptance view*. Cambridge: Cambridge University Press.
- Winch P (1958). *The idea of a social science*. London: Routledge & Kegan Paul.

Speech Acts

Y Huang, University of Reading, Reading, UK

© 2006 Elsevier Ltd. All rights reserved.

Speech act theory, though foreshadowed by the Austrian philosopher Ludwig Wittgenstein's views about language-games, is usually attributed to the Oxford philosopher J. L. Austin. The basic ideas, which were formed by him in the late 1930s, were presented in his lectures given at Oxford in 1952–1954, and later in his William James Lectures delivered at Harvard in 1955. These lectures were finally published posthumously as *How to do things with words* in 1962. After his death in 1960, Austin's ideas were refined, systematized, and advanced, especially by his Oxford pupil, the American philosopher John R. Searle. Simply stated, the central tenet of speech act theory is that the uttering of a sentence is, or is part of, an action within the framework of social institutions and conventions. Put in slogan form, saying is (part of) doing, or words are (part of) deeds.

J. L. Austin

The Performative/Constative Dichotomy

In the 1930s, a very influential school of thought in philosophy was logical positivism, developed by a group of philosophers and mathematicians principally in Vienna. One of the central doctrines of logical positivism is what is now called 'the descriptive fallacy,' namely, the view that the only philosophically interesting function of language is that of making true or false statements. A particular version of the descriptive fallacy is the verificationist thesis of

meaning, namely, the idea that 'unless a sentence can, at least in principle, be verified (i.e., tested for its truth or falsity), it was strictly speaking meaningless' (Levinson, 1983: 227). On such a view, sentences that are not used to make verifiable or falsifiable propositions are simply meaningless.

It was against this philosophical background that Austin set about to develop his theory of speech acts (Austin, 1962). He made two important observations. First, he noted that some ordinary language sentences such as those in (1) are not employed to make a statement, and as such they cannot be said to be true or false.

- (1a) Good afternoon!
- (1b) Is he a Republican?
- (1c) Come in, please.

Secondly and more importantly, Austin observed that there are ordinary language declarative sentences that similarly resist a truth-conditional analysis. The point of uttering such sentences is not just to say things, but also actively to do things. In other words, such utterances have both a descriptive and an effective aspect. Accordingly, Austin called them 'performatives,' and he distinguished them from assertions, or statement-making utterances, which he called 'constatives.' In other words, performatives are utterances that are used to do things or perform acts, as in (2), whereas constatives are utterances that are employed to make assertions or statements, as in (3).

- (2a) I christen/name this ship the Princess Elizabeth.
- (2b) I now pronounce you man/husband and wife.
- (2c) I promise to come to your talk tomorrow afternoon.
- (3a) My daughter is called Elizabeth.

- (3b) A freshly baked loaf doesn't cut easily.
- (3c) Maurice Garin won the first Tour de France in 1903.

Unlike those in (3), the declarative sentences in (2) have two characteristics: (i) they are not used intentionally to say anything, true or false, about states of affairs in the outside world, and (ii) their use constitutes (part of) an action, viz., that of christening/naming a ship in (2a), that of pronouncing a couple married in (2b), and that of promising in (2c). In addition, there are two further differences between (2a–b) and (2c). The first is that while (2a–b) is part of a conventional or ritual behavior supported by institutional facts (see also Strawson, 1964), (2c) is not. Secondly, while the performative verb, that is, the verb naming the action while performing it in (2a–b) is in general an essential element and cannot be omitted, it can in (2c). In other words, whereas, for example, we cannot christen/name a ship without using the verb *christen* or *name*, we can make a promise without using the verb *promise*, as in (4).

- (4) I'll come to your talk tomorrow afternoon.

Performatives can further be divided into two types: explicit and implicit. Explicit performatives are performative utterances that contain a performative verb that makes explicit what kind of act is being performed. By contrast, implicit performatives are performative utterances in which there is no such verb. Thus, the performatives in (2) are explicit ones, and the performative in (4) is an implicit one.

Austin also isolated a number of syntactic and semantic properties of explicit performatives in English: (i) explicit performatives contain a performative verb, (ii) the performative nature of such a verb can be reinforced by adding the adverb *hereby*, and (iii) explicit performatives occur in sentences with a first-person singular subject of a predicate (verb) in the simple present tense, indicative mood, and active voice.

However, as Austin himself was aware, there are exceptions. Explicit performatives can sometimes take a first-person plural subject, as in (5); a second-person singular or plural subject, as in (6); and a third-person singular or plural subject, as in (7). In addition, there are cases where the explicit performative verb is 'impersonal,' that is, it does not refer to the speaker, as in (8). Furthermore, as (6), (7), and (8) show, explicit performatives can occur in sentences where the verb is in the passive voice. Finally, as (9) indicates, they can also occur in sentences of present progressive aspect.

- (5) We suggest that you give up smoking immediately.

- (6) You are fired.
- (7) Passengers are hereby requested to wear a seat belt.
- (8) Notice is hereby given that shoplifters will be prosecuted.
- (9) I am warning you not to dance on the table.

Austin's Felicity Conditions on Performatives

As already mentioned, it makes no sense to call a performative true or false. Nevertheless, Austin noticed that for a performative to be successful or 'felicitous,' it must meet a set of conditions. For example, one such condition for the speech act of naming is that the speaker be recognized by his or her community as having the authority to perform that act; for the speech act of ordering, the condition is that the speaker have authority over the addressee; and finally, for the speech act of promising, one condition is that what is promised by the speaker must be something the addressee wants to happen. Austin called these conditions 'felicity conditions,' of which he distinguished three types, as shown in (10).

- (10) Austin's felicity conditions on performatives
 - (10a) (i) There must be a conventional procedure having a conventional effect.
 - (10a) (ii) The circumstances and persons must be appropriate, as specified in the procedure.
 - (10b) The procedure must be executed (i) correctly and (ii) completely.
 - (10c) Often
 - (i) the persons must have the requisite thoughts, feelings and intentions, as specified in the procedure, and
 - (ii) if consequent conduct is specified, then the relevant parties must so do.

Violation of any of the conditions in (10) will render a performative 'unhappy' or infelicitous. If conditions *a* or *b* are not observed, then what Austin described as a 'misfire' takes place. For instance, in England, a registrar conducting a marriage ceremony in an unauthorized place will violate condition *a* (i), thus committing a misfire. The same is true for a clergyman baptizing the wrong baby, because in this case, condition *a* (ii) is not fulfilled. Next, as an illustration of a violation of condition *b* (i), consider the case of a bridegroom not saying the exact words that are conventionally laid down for a Church of England marriage ceremony. As to condition *b* (ii), it dictates that the procedure be complete. Thus, in making a bet, the bet is not 'on' unless *You are on* (or something with the same effect) is uttered by the addressee; in Austin's terminology, this counts as a satisfactory 'uptake,' the absence of which will again

cause a misfire. Finally, if condition *c* is violated, then what Austin called an ‘abuse’ is committed (including, but not only, cases of insincerity). Examples include: congratulating someone when one knows that he or she passed his or her examination by cheating (condition *c* (i)); making a promise when one already intends to break it (condition *c* (ii)); and marrying without intending to consummate the marriage (see also Sadock’s (2004) discussion of these conditions in terms of misinvocation, misexecution, and abuse).

Locutionary, Illocutionary, and Perlocutionary Speech Acts

The initial distinction made by Austin between performatives and constatives was soon to be rejected by him in favor of a general theory of speech acts. In fact, as pointed by Levinson (1983: 231), there are two internal shifts in Austin’s arguments. First, there is a shift from the view that performatives are a special class of sentences/utterances with peculiar syntactic and semantic properties to the view that there is a general class of performatives that encompasses both explicit and implicit performatives, the latter including many other types of sentences/utterances. The second shift is from the performative/constative dichotomy to a general theory of speech acts, of which the various performatives and constatives are just special sub-cases.

What led Austin to abandon the performative/constative dichotomy? In the first place, he noted that like performatives, constatives are also subject to the felicity conditions stated in (10). Consider the so-called ‘Moore’s paradox,’ illustrated by (11).

- (11) ?Princess Diana died in a fatal car crash in Paris with Dodi Al Fayed, but I don’t believe it.

This utterance is infelicitous because it violates condition *c* (i) in (10) above. In the same vein, if someone utters (12) when he or she knows that John does not in fact have a wife, then its presupposition will not go through (see **Presupposition**). The reason the presupposition fails to carry through is that condition *a* (ii) in (10) above is not adhered to.

- (12) I’m sure John’s wife is a feminist.

Secondly, Austin observed that performatives and constatives may be impossible to distinguish even in truth-conditional terms. On the one hand, there are ‘loose’ constatives that may not be assessed strictly by means of truth conditions, as in (13). On the other hand, there are utterances like those in (14) that pass the *hereby* test and therefore are performatives by definition but that nevertheless are used to state or assert. In these cases, the performatives must be

counted simultaneously as constatives. On the basis of such evidence, Austin concluded that constatives are nothing but a special class of performatives, and that the two-way distinction between performatives, as action-performers, and constatives, as truth-bearers, can no longer be maintained.

- (13a) France is hexagonal.
 (13b) The fridge is empty.
 (13c) New York is sixty miles from where I live.
 (14a) I hereby state that Da Vinci started to paint *Mona Lisa* in 1503.
 (14b) I hereby tell you that the bill is right.
 (14c) I hereby hypothesize that there is water on Mars.

Consequently, Austin claimed that all utterances, in addition to meaning whatever they mean, perform specific acts via the specific communicative force of an utterance. Furthermore, he introduced a three-fold distinction among the acts one simultaneously performs when saying something, as illustrated in (15):

- (15) (A speech act’s three facets)
 (i) Locutionary act: the production of a meaningful linguistic expression.
 (ii) Illocutionary act: the action intended to be performed by a speaker in uttering a linguistic expression, by virtue of the conventional force associated with it, either explicitly or implicitly.
 (iii) Perlocutionary act: the bringing about of consequences or effects on the audience through the uttering of a linguistic expression, such consequences or effects being special to the circumstances of utterance.

A locutionary act is the basic act of speaking, which itself consists of three related sub-acts: (i) a phonic act of producing an utterance-inscription; (ii) a phatic act of composing a particular linguistic expression in a particular language; and (iii) a rhetic act of contextualizing the utterance-inscription (Austin, 1962). The first of these three sub-acts is concerned with the physical act of producing a certain sequence of vocal sounds (in the case of spoken language), or a set of written symbols (in the case of written language). The second refers to the act of constructing a well-formed string of sounds/symbols (a word, phrase, or sentence in a particular language). The third sub-act is responsible for tasks such as assigning reference, resolving deixis, and disambiguating the utterance-inscription lexically and/or grammatically (see **Deixis and Anaphora: Pragmatic Approaches**).

The illocutionary act refers to the fact that when we say something, we usually say it with some purpose in mind. In other words, an illocutionary act refers to the type of function the speaker intends to fulfill, or

the action the speaker intends to accomplish in the course of producing an utterance; it is also an act defined within a system of social conventions. In short, it is an act accomplished in speaking. Examples of illocutionary acts include accusing, apologizing, blaming, congratulating, declaring war, giving permission, joking, marrying, nagging, naming, promising, ordering, refusing, swearing, and thanking. The functions or actions just mentioned are also commonly referred to as the illocutionary 'force' (or 'point') of the utterance. Illocutionary force is frequently conveyed by what Searle (1969) called an 'illocutionary force indicating device' (IFID), the most direct and conventional type of which is an explicit performative in the form of (16) (where Vp stands for performative verb). Indeed, the term 'speech act' in its narrow sense is often taken to refer exclusively to illocutionary acts.

(16) I (hereby) Vp you (that) S

It should be mentioned at this point that the same linguistic expression can be used to carry out a wide variety of different speech acts, so that the same locutionary act can count as having different illocutionary forces in different contexts. Depending on the circumstances, one may utter (17) to make a threat, to issue a warning, or to give an explanation.

(17) The gun is loaded.

In fact, Alston (1994) has argued that the meaning of a sentence consists in its having a certain illocutionary act potential (IAP) that is closely and conventionally associated with its form. On this view, to know what a sentence means is to know what range of illocutionary acts it can be conventionally used to perform.

Conversely, the same speech act can be performed by different linguistic expressions, or the same illocutionary force can be realized by means of different locutionary acts. The utterances in (18), for example, illustrate different ways of carrying out the same speech act of requesting.

(18) (At ticket office in railway station)

(18a) A day return ticket to
Oxford, please.

(18b) Can I have a day return ticket to
Oxford, please?

(18c) I'd like a day return ticket to Oxford.

Finally, a perlocutionary act concerns the effect an utterance may have on the addressee. Put slightly more technically, a perlocution is the act by which the illocution produces a certain effect in or exerts a certain influence on the addressee. Still another way to put it is that a perlocutionary act represents a consequence or by-product of speaking, whether intentional or not.

The effect of the act being performed by speaking is generally known as the perlocutionary effect. There is an extensive literature on the differentiation between locutionary, illocutionary, and perlocutionary acts (see e.g., Sadock (2004) for a survey).

J. R. Searle

Searle's Felicity Conditions on Speech Acts

Just as its truth conditions must be met by the world for a sentence to be said to be true, its felicity conditions must also be fulfilled by the world for a speech act to be said to be felicitous. Searle (1969) took the view that the felicity conditions put forward by Austin are not only ways in which a speech act can be appropriate or inappropriate, but that they also jointly constitute the illocutionary force. Put in a different way, the felicity conditions are the constitutive rules – rules that create the activity itself – of speech acts. On Searle's view, to perform a speech act is to obey certain conventional rules that are constitutive of that type of act. Searle developed the original Austinian felicity conditions into a neo-Austinian classification of four basic categories, namely (i) propositional content, (ii) preparatory condition, (iii) sincerity condition, and (iv) essential condition. As an illustration of these conditions, consider (19).

- (19) Searle's felicity conditions for promising
- (i) propositional content: future act A of S
 - (ii) preparatory: (a) H would prefer S's doing A to his not doing A, and S so believes
 - (b) It is not obvious to both S and H that S will do A in the normal course of events
 - (iii) sincerity: S intends to do A
 - (iv) essential: the utterance of e counts as an undertaking to do A

where S stands for the speaker, H for the hearer, A for the action, and e for the linguistic expression.

The propositional content condition is in essence concerned with what the speech act is about. That is, it has to do with specifying the restrictions on the content of what remains as the 'core' of the utterance (i.e., Searle's propositional act) after the illocutionary act part is removed. For a promise, the propositional content is to predicate some future act of the speaker, whereas the preparatory conditions state the real-world prerequisites for the speech act. In the case of a promise, the latter are roughly that the addressee would prefer the promised action to be accomplished, that the speaker knows this, but also that it is clear to both the speaker and the addressee that what is promised will not happen in the normal course of action. Next, the sincerity condition must be satisfied if the act is to be performed sincerely. When carrying out

an act of promising, the speaker must genuinely intend to keep the promise. Notice that if the sincerity condition is not fulfilled, the act is still performed, but there is an abuse, to use Austin's term. Finally, the essential condition defines the act being performed in the sense that the speaker has the intention that his or her utterance will count as an act, and that this intention is recognized by the addressee. Thus for a promise, the speaker must have the intention to create an obligation to act. Failure to meet the essential condition has the consequence that the act has not been carried out.

Searle's Typology of Speech Acts

Can speech acts be classified, and if so, how? Austin (1962) grouped them into five types: (i) verdictives: giving a verdict, (ii) exercitives: exercising power, rights or influence, (iii) commissives: promising or otherwise undertaking, (iv) behabitives: showing attitudes and social behavior, and (v) expositives: fitting an utterance into the course of an argument or conversation. Since then, there have been many attempts to systematize, strengthen, and develop the original Austinian taxonomy (Bach and Harnish, 1979; Allan, 2001; Bach, 2004). Some of these new classifications are formulated in formal/grammatical terms, others, in semantic/pragmatic terms, and still others, on the basis of the combined formal/grammatical and semantic/pragmatic modes (see Sadock (2004) for a review). Of all these (older and newer) schemes, Searle's (1975a) neo-Austinian typology remains the most influential. Under Searle's taxonomy, speech acts are universally grouped into five types along four dimensions: (i) illocutionary point, (ii) direction of fit between words and world, (iii) expressed psychological state, and (iv) propositional content (see also Searle (2002)). The five types of speech acts are further explained next.

(i) Representatives (or assertives; the constatives of the original Austinian performative/constative dichotomy) are those kinds of speech acts that commit the speaker to the truth of the expressed proposition and thus carry a truth-value. They express the speaker's belief. Paradigmatic cases include asserting, claiming, concluding, reporting, and stating. In performing this type of speech act, the speaker represents the world as he or she believes it is, thus making the words fit the world of belief.

(20) The Berlin Wall came down in 1989.

(ii) Directives are those kinds of speech acts that represent attempts by the speaker to get the addressee to do something. They express the speaker's desire/wish for the addressee to do something. Paradigmatic

cases include advice, commands, orders, questions, and requests. In using a directive, the speaker intends to elicit some future course of action on the part of the addressee, thus making the world match the words via the addressee.

(21) Put the cake in the oven.

(iii) Commissives are those kinds of speech acts that commit the speaker to some future course of action. They express the speaker's intention to do something. Paradigmatic cases include offers, pledges, promises, refusals, and threats. In the case of a commissive, the world is adapted to the words via the speaker himself.

(22) I'll never buy you another computer game.

(iv) Expressives are those kinds of speech acts that express a psychological attitude or state of the speaker such as joy, sorrow, and likes/dislikes. Paradigmatic cases include apologizing, blaming, congratulating, praising, and thanking. There is no direction of fit for this type of speech act.

(23) Well done, Elizabeth!

(v) Declarations (or declaratives) are those kinds of speech acts that effect immediate changes in some current state of affairs. Because they tend to rely on elaborate extralinguistic institutions for their successful performance, they may be called institutionalized performatives. In performing this type of speech act, the speaker brings about changes in the world; that is, he or she effects a correspondence between the propositional content and the world. Paradigmatic cases include (officially) opening a bridge, declaring war, excommunicating, firing from employment, and nominating a candidate. As to the direction of fit, it is both words-to-world and world-to-words.

(24) I object, Your Honor.

Illocutional point (or speech act type), direction of fit, and expressed psychological state can be summarized as in (25).

(25) Illocutionary point/Speech act type	Direction of fit	Expressed psychological state
representative	words-to-world	belief
directive	world-to-words	desire
commissive	world-to-words	intension
expressive	none	variable
declaration	both	none

Indirect Speech Acts

What is an indirect speech act? Most of the world's languages have three basic sentence types: declarative, interrogative, and imperative. In some languages,

the three major sentence types are distinguished morphologically and/or syntactically; as instances, compare Somali, Greenlandic, or Lakhota (Lakota) (see Huang (2006) for further discussion). The three sentence types are typically associated with the three basic illocutionary forces, namely, asserting/stating, asking/questioning, and ordering/requesting, respectively.

In the case of a direct match between a sentence type and an illocutionary force, we have a direct speech act. In addition, explicit performatives, which happen to be in the declarative form, are also taken to be direct speech acts, because they have their illocutionary force explicitly named by the performative verb in the main part (or 'matrix clause') of the sentence. On the other hand, if there is no direct relationship between a sentence type and an illocutionary force, we are faced with an indirect speech act. Thus, when an explicit performative is used to make a request, as in (26), it functions as a direct speech act; the same is the case when an imperative is employed, as in (27). By comparison, when an interrogative is used to make a request, as in (28), we have an indirect speech act.

(26) I request you to pass the salt.

(27) Pass the salt.

(28) Can you pass the salt?

In short, the validity of the distinction between direct and indirect speech acts is dependent upon whether or not one subscribes to what Levinson (1983: 264, 274) has called the 'literal force hypothesis' – the view that there is a direct structure-function correlation in speech acts and that sentence forms are direct reflexes of their underlying illocutionary forces.

There are, however, problems at the very heart of the literal force hypothesis. One is that there are cases of speech acts where even the direct link between performative verbs and speech acts breaks down. Consider (29).

(29) I promise to sack you if you don't finish the job by this weekend.

In (29), the performative verb is *promise*, but the force that is most naturally ascribed to this speech act is that of either a threat or a warning. This shows that, contrary to the literal force hypothesis, we cannot always identify speech acts, even with sentences containing a performative verb.

Secondly and more importantly, as also pointed out by Levinson (1983: 264), most usages are indirect. The speech act of requesting, for example, is very rarely performed by means of an imperative in English. Instead, it is standardly carried out indirectly.

Furthermore, there are probably infinitely many varieties of sentences that can be used to indirectly make a request, as shown in (30).

(30a) I want you to put the cake in the oven.

(30b) Can you put the cake in the oven?

(30c) Will you put the cake in the oven?

(30d) Would you put the cake in the oven?

(30e) Would you mind putting the cake in the oven?

(30f) You ought to put the cake in the oven.

(30g) May I ask you to put the cake in the oven?

(30h) I wonder if you'd mind putting the cake in the oven?

As to how to analyze indirect speech acts, there are roughly three approaches. The first is to assume the existence of a dual illocutionary force (as proposed by Searle, 1975b). On this assumption, indirect speech acts have two illocutionary forces, one literal or direct, and the other nonliteral or indirect. While the literal force is secondary, the nonliteral force is primary. Next, whether an utterance operates as an indirect speech act or not has to do with the relevant felicity conditions. For example, (28) both infringes the felicity condition for a question and queries the preparatory condition for a request. This explains why it can function as an indirect speech act, whereas (31), for instance, cannot; the reason is that in the case of (31), felicity conditions are irrelevant.

(31) Salt is made of sodium chloride.

Finally, on Searle's view, because a speaker's performing and an addressee's understanding an indirect speech act always involves some kind of inference, the question is how this inference can be computed. Searle's suggestion is that it can be computed along the general lines of the rational, cooperative model of communication articulated by Grice (1989) (see **Cooperative Principle**).

One interesting characteristic of indirect speech acts is that they are frequently conventionalized (see, e.g., Morgan, 1978). This can be illustrated by the fact that of various, apparently synonymous linguistic expressions, only one may conventionally be used to convey an indirect speech act, as illustrated in (32).

(32a) Are you able to pass the salt?

(32b) Do you have the ability to pass the salt?

Under Searle's analysis, both (32a) and (32b) would be expected to be able to perform the indirect speech act of requesting, because (i) they are largely synonymous with (28), and (ii) they, too, inquire about the satisfaction of the addressee-based preparatory condition for making a request. But this expectation is not fulfilled.

Searle's response to this puzzle is that there is also a certain degree of conventionality about indirect speech acts, and that this may be accounted for in terms of conventions of meaning/usage. Inspired by this insight, Morgan (1978) developed the notion of 'short-circuited implicature' to cover inference involved in cases like (28) (see **Implicature**). While the relevant implicature is in principle calculable, in practice it is not calculated in cases like these. From a linguistic point of view, the conventionality here is correlated with the possible occurrence of *please*. While *please* can be inserted before the verb *pass* in (26)–(28), it cannot in (32), as shown in (33).

- (33a) I request you to please pass the salt.
- (33b) Please pass the salt.
- (33c) Can you please pass the salt?
- (33d) ?Are you able to please pass the salt?
- (33e) ?Do you have the ability to please pass the salt?

Furthermore, the conventionality indicated by *please* in (33a) and (33b) is one of meaning, hence the speech act of requesting is performed directly. By contrast, the conventionality signaled by *please* in (33c) is one of usage, and thus we have an indirect speech act.

A second, rather similar, approach is due to Gordon and Lakoff (1975). In their analysis, there are inference rules called 'conversational postulates' that reduce the amount of inference needed to interpret an indirect speech act. Thus, in the case of (28), if the interpretation as a question cannot be intended by the speaker, then the utterance will be read as being equivalent to his or her having said (26), thus resulting in the performance of the indirect speech act of requesting. Stated this way, the conversational postulates proposed by Gordon and Lakoff can be seen as another reflection of the conventionality of indirect speech acts. As to the similarities and differences between Searle's and Gordon and Lakoff's analyses, the major similarity is that both accounts assume that the interpretation of indirect speech acts involves inference as well as conventionality; the major difference concerns the question of balance, namely, how much of the work involved in computing an indirect speech act is inferential and how much is conventional.

Finally, in contrast to the inferential models we have just discussed, there is the idiom model. In this model, sentences like (28) are semantically ambiguous, and the request interpretation constitutes a speech act idiom that involves no inference at all. On this view, (28) is simply recognized as a request, with no question being perceived. This is the position taken by Sadock (1974). There are, however, problems with this analysis, too. One is that it fails to

capture the fact that (in contrast to what is the case for idioms) the meaning of an indirect speech act can frequently (at least in part) be derived from the meaning of its components (the technical term for this is 'compositionality'); in addition, these would-be 'idioms' turn out to be quite comparable cross-linguistically (something which idioms are not). For example, an utterance like (34) may be used, with the same force as in English, in its Arabic, Chinese, German, or Modern Greek versions to indirectly request the addressee to switch on the central heating system (of course, always depending on the context).

- (34) 'It's cold in here.'

A further problem is that in the idiom model, an interpretation that takes into account the literal meaning or the direct illocutionary force of an indirect speech act is not allowed. This, however, leaves examples like (35) unexplained.

- (35) A: Can you pass the salt?
B: Yes, I can. (Here it is.)

Why, then, do people use indirect speech acts? One answer is that the use of indirect speech acts is in general associated with politeness. Indirect speech acts are usually considered to be more polite than their direct counterparts (see the considerable literature on the analysis of speech acts, especially the work on face-threatening acts (FTAs) like requests, complaints, and apologies in the tradition of Brown and Levinson's (1987) classical 'face-saving' politeness model).

Speech Acts and Culture

Cross-Cultural Variation

Many speech acts are culture-specific. This is particularly so in the case of institutionalized speech acts, which typically use standardized and stereotyped formulae and are performed in public ceremonies. A good example is provided by the speech act of divorcing. In some Muslim cultures, under the appropriate circumstances, the uttering of a sentence with the import of (36) three times consecutively by a husband to his wife will *ipso facto* constitute a divorce. By contrast, in Western cultures, no one (no matter what his or her religion is) can felicitously use (36) to obtain a divorce.

- (36) 'I hereby divorce you.'

But how about non-institutionalized speech acts? First of all, as said above, any given speech act may be culture-specific. Rosaldo (1982), for example, observed that the speech act of promising has no place among the Ilongots – a tribal group of hunters

and horticulturalists in the Philippines. She attributes the absence of this speech act in the conceptual repertoire of the Ilongot to a lack of interest in sincerity and truth in that community. The Ilongot, argues Rosaldo, are more concerned with social relationships than with personal intentions. On the basis of anthropological evidence such as this, Rosaldo claims that the universality of Searle's typology of speech acts cannot be maintained. Another example of this kind has been reported for the Australian aboriginal language Yolngu. According to Harris (1984: 134–135), there does not seem to be any speech act of thanking in the Yolngu speaker's repertoire.

Conversely, a given speech act may be present only in certain cultures. For example, in the Australian aboriginal language Walmajari, one finds a speech act of requesting that is based on kinship rights and obligations. The verb in question is *japirlyung* (Hudson, 1985), and the speech act may be called 'kinship-based requesting,' because it conveys a message meaning roughly 'I ask/request you to do X for me, and I expect you to do it simply because of how you are related to me'. Thus, for the speakers of Walmajari, it is very difficult to refuse a kinship-based speech act of requesting (see also Wierzbicka, 1991: 159–160). 'Exotic' speech acts such as the kinship-based requesting do not seem to be present in other East Asian or Western cultures.

Secondly, given a particular situation, pertinent speech acts are carried out differently in different cultures. For instance, in some East Asian and Western cultures, if one steps on another person's toes, one normally performs the speech act of apologizing. But apparently this is not the case among the Akans, a West African culture. As reported by Mey (2001: 287, crediting Felix Ameka), in that culture, such a situation does not call for apologies but calls for the expression of sympathy: "The focus is on the person to whom the bad thing has happened rather than the person who has caused the bad thing" (Mey, 2001: 287). Another example: while in English, thanks and compliments are usually offered to the hosts when leaving a dinner party, in Japanese society, apologies such as *o-jama itashimashita* 'I have intruded on you' are more likely to be offered by the guests. A similar speech act of apologizing is performed in Japanese upon receiving a present, when a Japanese speaker is likely to say something like *sumimasen* – the most common Japanese 'apology' formula or one of its variants. Conversely (as pointed out by many authors), apologies can be used in a much broader range of speech situations in Japanese than in English.

Thirdly, in different cultures/languages, the same speech act may meet with different typical responses.

For example, a compliment normally generates acceptance/thanking in English, but self-denigration in Chinese, Japanese, or even Polish. A typical compliment/response formula in Chinese would be something like (37).

- (37) A: *ni cai zuode zhen hao!*
 B: *nali, nali, wo bu hui zuocai.*
 A: *bie keqi. ni cai zhende zuode hen hao!*
 B: *ni tai keqi le.*
 A: 'You cook really well!'
 B: 'No, no, I don't really know how to cook properly.'
 A: 'Please don't be too modest. You really cook very well.'
 B: 'You're too kind.'

The same is even more true in Japanese. According to Mizutani and Mizutani (1987: 43), "[T]he Japanese will never accept a compliment without saying *ie* ['no']". Given the general Japanese reluctance to say 'no' under almost any other circumstances, the compliment response pattern is rather striking.

Fourthly, the same speech act may differ in its directness/indirectness in different cultures. Since the late 1970s, a great deal of research has been conducted on how particular kinds of speech acts, especially such face-threatening acts as requests, apologies, and complaints are realized across different languages. Of these investigations, the most influential is the large-scale Cross-Cultural Speech Act Realization Patterns Project (CCSARP) (Blum-Kulka *et al.*, 1989). In this project, the realization patterns of requesting and apologizing in German; Hebrew; Danish; Canadian French; Argentinean Spanish; and British, American, and Australian English were compared and contrasted. In the case of requests, the findings were that among the languages examined, the Argentinean Spanish speakers are the most direct, followed by the speakers of Hebrew. The least direct are the Australian English speakers, while the speakers of Canadian French and German are positioned at the midpoint of the directness/indirectness continuum. Building on the CCSARP, strategies for the performance of certain types of face-threatening acts in a much wider range of languages have since been examined. These languages include Catalan, Chinese, Danish, Dutch, French, German, Greek, Hebrew, Japanese, Javanese, Polish, Russian, Thai, Turkish, four varieties of English (British, American, Australian, and New Zealand), two varieties of French (Canadian and French), and eight varieties of Spanish (Argentinean, Ecuadorian, Mexican, Peninsular, Peruvian, Puerto Rican, Uruguayan, and Venezuelan). As a result of these studies, it has now been established that there is indeed

extensive cross-cultural variation in directness/indirectness in speech acting, especially in the realization of face-threatening acts (FTAs), and that these differences are generally associated with the different means that different languages utilize to realize speech acts. These findings have undoubtedly contributed to our better understanding of cross-cultural/linguistic similarities and differences in face-redressive strategies for FTAs (see Huang (2006) for detailed discussion).

Interlanguage Variation

A number of studies have recently appeared that explore speech acts in interlanguage pragmatics. Simply put, an interlanguage is a stage on a continuum within a rule-governed language system that is developed by L2 learners on the way to acquiring the target language. This language system is intermediate between the learner's native language and his or her target language.

Some of these studies investigate how a particular type of speech act is performed by non-native speakers in a given interlanguage; others compare and contrast the similarities and differences in the realization patterns of given speech acts between native and non-native speakers of a particular language. The best studied interlanguage is that developed by speakers of English as a second language. Other interlanguages that have been investigated include Chinese, German, Hebrew, Japanese, and Spanish (see Huang (2006) for further discussion).

A few recent formal and computational approaches to speech acts and speech act theory are worthy of note. One important theoretical development is the integration of speech acts with intensional logic, resulting in what is called 'illocutionary logic' (Searle and Vanderveken, 1985; Vanderveken, 1994, 2002). Similarly, Merin (1994) has endeavored to produce an algebra of what he calls 'social acts.' Finally, recent formalizations of various aspects of speech act theory in artificial intelligence and computational linguistics can be found in Perrault (1990), Bunt and Black (2000), and Jurafsky (2004) (see also Sadock, 2004).

See also: Cooperative Principle; Deixis and Anaphora; Pragmatic Approaches; Implicature; Mood, Clause Types, and Illocutionary Force; Presupposition.

Bibliography

- Allan K (2001). *Natural language semantics*. Oxford: Blackwell.
- Alston W (1994). 'Illocutionary acts and linguistic meaning.' In Tsohatzidis S L (ed.) *Foundations of speech*

- act theory: philosophical and linguistic perspectives*. London: Routledge. 29–49.
- Austin J L (1962). *How to do things with words*. Oxford: Oxford University Press.
- Bach K (2004). 'Pragmatics and the philosophy of language.' In Horn L & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 463–488.
- Bach K & Harnish R M (1979). *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.
- Blum-Kulka S, House J & Kasper G (eds.) (1989). *Cross-cultural pragmatics: requests and apologies*. Norwood: Ablex.
- Brown P & Levinson S C (1987). *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
- Bunt H & Black W (eds.) (2000). *Abduction, belief and context in dialogue: studies in computational pragmatics*. Amsterdam: John Benjamins.
- Gordon D & Lakoff G (1975). 'Conversational postulates.' In Cole P & Morgan J L (eds.) *Syntax and semantics 3: speech acts*. New York: Academic Press. 83–106.
- Grice H P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Harris S G (1984). *Culture and learning: tradition and education in north-east Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.
- Huang Y (2006). *Pragmatics*. Oxford: Oxford University Press.
- Hudson J (1985). 'Selected speech act verbs in Walmatjari.' In Huttar G & Gregerson K (eds.) *Pragmatics in non-Western practice*. Dallas: Summer Institute of Linguistics. 63–83.
- Jurafsky D (2004). 'Pragmatics and computational linguistics.' In Horn L & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 578–604.
- Levinson S C (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Merin A (1994). 'Algebra of elementary social acts.' In Tsohatzidis S L (ed.) *Foundations of speech act theory: philosophical and linguistic perspectives*. London: Routledge. 234–266.
- Mey J L (2001). *Pragmatics: an introduction* (2nd edn.). Oxford: Blackwell.
- Mizutani O & Mizutani N (1987). *How to be polite in Japanese*. Tokyo: Japan Times.
- Morgan J L (1978). 'Two types of convention in indirect speech acts.' In Cole P (ed.) *Syntax and semantics 9: Pragmatics*. New York: Academic Press. 261–280.
- Perrault C R (1990). 'An application of default logic to speech act theory.' In Cohen P R, Morgan J & Pollack M E (eds.) *Intentions in communication*. Cambridge, MA: MIT Press. 161–186.
- Rosaldo M Z (1982). 'The things we do with words: Ilongot speech acts and speech act theory in philosophy.' *Language in Society* 11, 203–237.
- Sadock J (1974). *Toward a linguistic theory of speech acts*. New York: Academic Press.
- Sadock J (2004). 'Speech acts.' In Horn L & Ward G (eds.) *The handbook of pragmatics*. Oxford: Blackwell. 53–73.

- Searle J (1969). *Speech acts*. Cambridge: Cambridge University Press.
- Searle J (1975a). 'A taxonomy of Speech acts.' In Gunderson K (ed.) *Minnesota studies in the philosophy of science 9: Language, mind and knowledge*. 344–369.
- Searle J R (1975b). 'Indirect speech acts.' In Cole P & Morgan J L (eds.) *Syntax and semantics 3: Speech acts*. New York: Academic Press. 59–82.
- Searle J R (2002). *Consciousness and language*. Cambridge: Cambridge University Press.
- Searle J R & Vanderveken D (1985). *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.
- Strawson P F (1964). 'Intention and convention in speech acts.' *Philosophical Review* 73, 290–302.
- Vanderveken D (1994). 'A complete formulation of a simple logic of elementary illocutionary acts.' In Tsohatzidis S L (ed.) *Foundations of speech act theory: philosophical and linguistic perspectives*. London: Routledge. 99–131.
- Vanderveken D (2002). 'Universal grammar and speech act theory.' In Vanderveken D & Kubo S (eds.) *Essays in speech act theory*. Amsterdam: John Benjamins. 25–62.
- Wierzbicka A (1991). *Cross-cultural pragmatics: the semantics of human interaction*. Berlin: Mouton de Gruyter.

Syncategoremata

G Klima, Fordham University, Bronx, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

The History of the Distinction

The medieval distinction between syncategorematic terms ('syncategoremata,' such as 'and,' 'or,' and 'not,' which merely combine other terms) and categorematic terms ('categoremata,' such as 'man,' 'Socrates,' and 'animal,' which can be predicated of things) dates back to the 6th-century grammarian Priscian, who in his *Institutiones Grammaticae*, who in his turn, attributes the idea to certain 'dialecticians.' Syncategoremata were commonly recognized by Stoics and Neo-Platonics (Pinborg, 1972: 60). However, Peripatetic dialecticians could also take their cue from Aristotle's relevant remark on the copula (the verb 'is' or its equivalent, joining subject and predicate), when he says that it is in itself "nothing, but co-signifies some combination, which cannot be thought of without the components" (*On Interpretation*, 16b24–25). In any case, according to Priscian, the dialecticians he refers to characterized syncategoremata in a similar way: such terms merely cosignify, in contrast to verbs and nouns, which, when combined, make a complete expression. With the emergence of the distinctive logical and grammatical literature of the Middle Ages in the 12th century continuing through the late 15th and early 16th centuries, discussions of syncategoremata became part of the characteristic genres of this literature (Sweeney, 2002), in which our medieval colleagues developed ever more sophisticated characterizations of syncategorematic terms, their distinction from categorematic terms, their function, and their nature. According to Norman Kretzmann's (1982) periodization, we can

distinguish the following main stages in the medieval career of syncategoremata:

1. Their emergence as the focal points of certain logical or semantic relationships or special problems of interpretation (in the 12th century, especially the latter half)
2. Their identification as a distinguishable set of topics worthy of development in separate treatises typically called syncategoremata (from the last quarter of the 12th century to the last quarter of the 13th century)
3. Their assimilation into general treatises on logic, sometimes as a group but sometimes dispersed in ways designed to associate particular syncategoremata with more general topics in logic to which they are appropriate
4. Their absorption into the sophisma literature (discussions of logical puzzles generated by problem sentences, the so-called 'sophismata'), in which a particular syncategorema may serve as the germ of a paradox the interest of which is often associated with metaphysics or natural philosophy more than with logic or semantics proper (from the first quarter of the 14th century to the disintegration of scholastic logic).

With the decline of scholastic logic, the logical treatises specifically devoted to syncategoremata or to sophismata focusing on the problems generated by syncategoremata became gradually extinct along with the rest of the characteristic logical literature of the Middle Ages. Yet, the distinction between categorematic and syncategorematic terms has never become entirely forgotten, and it keeps recurring in various guises also in modern philosophy. However, the most sophisticated accounts of the distinction are

still to be found in the medieval literature, which is the focus of the rest of this article.

Syntactic and Semantic Criteria of Drawing the Distinction

It is interesting that Priscian's remark involved two different and by no means equivalent criteria for distinguishing syncategorematic and categorematic terms. On what may be called the 'syntactic' criterion, categorematic terms are those that can function as subjects or predicates of propositions so that their combination yields a complete proposition, whereas syncategorematic terms cannot be subjects or predicates. On the 'semantic' criterion, however, syncategorematic terms are those that merely cosignify (i.e., signify in combination with some other term) but that in themselves do not signify anything. Apparently, the syntactic criterion provides a good test for sorting various parts of speech into either of the two members of the division. However, even if the criterion works, it does not seem to provide a principled reason why certain parts of speech behave in one way or the other. The semantic criterion seems to provide such a reason, but as it stands, it is rather vague.

Indeed, on closer inspection even the syntactic criterion does not seem to fare much better. First, in the appropriate context any part of speech can be a subject or a predicate. For example, the preposition 'of' or the negation 'not,' which are certainly obvious candidates for being regarded as syncategorematic, can be used as subjects in the sentences '*Of* is a preposition' and '*Not* is a negation,' where they are taken to stand for themselves (and other token phrases of the same type) or, as medieval logicians would put it, in 'material supposition.' (Cf. the modern distinction between mention and use.) Accordingly, later medieval authors sometimes refined the syntactic criterion by adding the requirement that syncategorematic terms are those that cannot be subjects or predicates when used 'significatively' (i.e., in their proper function) and not with the intent to take them for themselves. However, this refinement clearly indicates that the distinction primarily applies not to words per se but, rather, to their different uses. This point is further reinforced by the fact that parts of speech that on the simple syntactic criterion would be deemed syncategorematic in some of their uses can be subjects or predicates even when taken significatively. For instance, adjectives that cannot serve as subjects on their own can be predicates (e.g., contrast the grammatical 'A man is brave' with the ungrammatical 'A brave is a man'), and they can have substantive uses (in Latin marked by the neuter gender),

in which they can even serve as subjects (as in the Latin sentence, *Album est coloratum* '[What is] white is colored,' but even in English we can say, 'Blue is a soothing color'). Furthermore, logicians, who were primarily interested in the semantic features of syncategoremata, would not regard adjectives and adverbs as syncategorematic without further ado.

Therefore, to distinguish such terms from syncategoremata pure and simple, they introduced a further distinction. The two most influential authors of the 13th century on this topic, William of Sherwood and Peter of Spain, both distinguished 'pure categoremata' (i.e., subjects and predicates) from their dispositions or determinations (Peter of Spain, 1992: 38–41; William of Sherwood, 1941: 48). However, they also distinguished those dispositions of subjects and predicates that belong to them insofar as they are subjects and predicates (such as 'signs of quantity,' i.e., quantifiers), the pure syncategoremata, and those dispositions that belong to the things that are signified by subjects and predicates regardless of the fact that they are subjects and predicates. The latter are just categorematic dispositions of pure categoremata, such as adjectives and adverbs. For example, in the proposition 'Every wise man is running,' the pure syncategorematic term 'every' is a determination of the subject term of this sentence in relation to the predicate, determining that the sentence is true only if the predicate applies to everything that falls under the subject. Thus, 'every' applies to the subject insofar as it is the subject of the predicate of this sentence. However, the determination 'wise' pertains to 'man' regardless of what the latter is subjected to and how. This determination merely determines that of all man only the wise are considered (i.e., those things that have the property signified by the adjective).

Thus, adjectives and adverbs, even in their purely adjectival or adverbial uses, on this criterion are no longer regarded as pure syncategoremata, even if they cannot be self-standing subjects or predicates. Rather, they are regarded as categorematic parts of complex subjects and predicates signifying the dispositions of the things signified by nouns and verbs, the principal parts of such complex subjects and predicates.

Further refinements of the distinction were provided by the nominalist philosophers of the 14th century (see **Nominalism**), such as William Ockham, John Buridan, and Albert of Saxony, with reference to their conception of a 'mental language,' a natural system of representation constituted by mental concepts, the mental acts of a human mind to which spoken or written parts of speech are systematically subordinated, rendering these spoken or written signs

meaningful (*see Mentalese*). In this setting, any part of a spoken or written language is said to signify immediately the concept to which it is subordinated, and it is said to signify ultimately the object or objects conceived by means of the concept to which it is subordinated (Buridan, 2001: xxxiv–xli). However, since some concepts have only the function of combining simple concepts into complex ones (e.g., a mental copula or conjunction), or just to modify the representative function of other concepts (e.g., a mental term-negation), but in themselves do not have the function of representing any objects, such concepts themselves are syncategorematic in mental language. Thus, the purely syncategorematic terms of spoken or written languages will be those that are subordinated to such syncategorematic concepts.

For example, as Buridan (2001: 234) remarks,

The copulas ‘is’ and ‘is not’ signify different ways of combining mental terms in order to form mental propositions, and *these different ways [of combining] are in their turn complexive concepts*. . . . And so also the words ‘and,’ ‘or,’ ‘if,’ ‘therefore,’ and the like designate *complexive concepts* that combine several propositions or terms at once in the mind, but nothing further outside the mind. These words are called purely syncategorematic, because they signify nothing outside the mind, except along with others, in the sense that the whole complex consisting of categorematic and syncategorematic words does signify the things conceived outside the mind, but this is on account of the categorematic words.

Buridan’s younger contemporary, Albert of Saxony (1974: f. 2v), provides further clarification of what we should understand by the significative function of syncategorematic terms when he observes:

If the terms ‘every,’ ‘no,’ etc. are taken materially, then they certainly can be subjects or predicates of propositions, as when we say ‘Every is a universal sign,’ or ‘Or is a [disjunctive] conjunction,’ or ‘No is an adverb,’ or ‘And is a copulative conjunction.’ For in these propositions these words are not taken significatively, since *they do not exercise the function [non exercent officium] which they were imposed to exercise*. So, in the proposition ‘Every is a universal sign,’ ‘every’ does not distribute anything, in ‘No is an adverb,’ ‘no’ does not negate anything, and likewise, in ‘And is a copulative conjunction,’ ‘and’ does not copulate anything; therefore, in these propositions these terms are taken not syncategorematically, but categorematically.

Therefore, syncategorematic terms, when they are taken significatively, are imposed to exercise the logical functions of modifying the semantic functions of categorematic terms with which they are construed on account of being subordinated to mental acts that exercise these functions on the mental level. Indeed, as Albert of Saxony (1974: f. 3r) notes, this is

precisely the reason why in their significative function they cannot be subjects of predicates. Thus, Albert presents the semantic distinction on the mental level as the ultimate reason for the syntactic criterion of the distinction on the spoken level. Indeed, later medieval treatises (in the late 14th and early 15th centuries) on concepts and mental language, such as those by Thomas of Cleves, Paul of Gelria, or Peter of Ailly, draw the distinction as pertaining directly and primarily to acts of the mind. These authors describe ‘mental categoremata’ as being acts of conceiving of some objects, as opposed to ‘mental syncategoremata,’ which are rather different modes or ways of conceiving of the objects conceived by the former (Bos and Read, 2001: 14, 54–57, 96, 130; Peter of Ailly, 1980: 18–19; see also the previous quote from Buridan). From this characterization, and from the common medieval doctrine that the semantic functions of spoken and written terms are determined by the semantic functions of the concepts to which they are subordinated, we get as a simple corollary the usual semantic and syntactic features of (pure) syncategoremata, namely that they are cosignificative rather than significative on their own, and that taken significatively they cannot be referring terms of propositions.

Although the distinction drawn in these terms was not in vogue in early modern philosophy, it is significant that John Locke would characterize the parts of speech he calls ‘particles’ (but that his medieval colleagues would recognize as syncategoremata) as being “marks of some action or intimation of the mind” – that is, some mental operation on ideas of the mind (Locke, 1995: bk. IV, c. 7, n. 4). Later, Immanuel Kant would describe pure concepts of understanding as ‘logical functions,’ directly preparing the way for Gottlob Frege to describe quantifiers as ‘second order concepts’ (i.e., as concepts operating on concepts).

Philosophical Significance of the Distinction

Syncategorematic terms present a particular problem for those philosophers who would take the primary function of elements of a language to be the signs or names of things. For in the case of syncategorematic terms, we just do not seem to have obvious candidates among things in the world for these terms to name. Nevertheless, medieval realists, such as Peter of Spain or William of Sherwood, or the modistic grammarians of the 13th and 14th centuries, such as Thomas of Erfurt (for further references, see Zupko, 2003), who professed a close parallelism between *modi significandi* (modes of signifying), *modi intelligendi*

(modes of understanding), and *modi essendi* (modes of being), assumed in their ontology certain modes of being or dispositions of things corresponding to the modes of signifying determined by modes of understanding conveyed by syncategorematic terms. Therefore, in this framework, syncategorematic terms do have counterparts in reality: the modes or dispositions of things that ‘prompt’ us to conceive of and hence signify things in certain ways.

Although these philosophers, logicians, and grammarians were careful to attach only a certain diminished degree of reality to these dispositions, nevertheless, it was precisely their ‘ontological liberalism’ that prompted William Ockham and his nominalist followers discussed previously to identify syncategoremata primarily with the mental acts modifying the representative function of categorematic concepts. For in this nominalist framework, syncategorematic terms simply have no extramental counterparts: the complex expressions they form with categorematic terms signify only the things signified by the categorematic terms, but in different ways. Thus, for example, as Buridan explicitly concluded, the proposition ‘God is God’ and the proposition ‘God is not God’ signify extramentally exactly the same thing as the term ‘God’ does, but this does not render these expressions synonymous; they signify the same thing differently because of the different syncategoremata signified by them in the mind (Buridan, 2001: 234). This is how late-medieval nominalists were able to have a parsimonious ontology along with a sufficiently ‘fine-grained’ semantics, by making the necessary semantic distinctions on the conceptual and not on the ontological level.

In addition to these and similar ontological considerations, the reinterpretation of the medieval distinction had even more far-reaching consequences in early modern and modern philosophy with Kant’s reclassification of a number of traditional metaphysical concepts, such as substance and accident, cause and effect, existence, necessity, and possibility, as concepts of pure understanding – that is, as logical functions or syncategoremata. Kant’s considerations, especially those concerning existence, directly paved the way for Gottlob Frege’s analysis of the notion of existence as a second-order concept, the existential quantifier, which in turn could immediately be exploited by the anti-metaphysical program of logical positivists, such as Rudolf Carnap (1959). Accordingly, without the recent reevaluation of the concepts of existence, possibility, and ontological commitment in the framework of modern possible-worlds

semantics, metaphysics probably still could not be regarded as the legitimate philosophical subject in analytic philosophy it has become in the past few decades. As even this example shows, our actual understanding of the medieval distinction still has fundamental significance in our considerations concerning the relationships between language, thought, and reality.

See also: Concepts; Mentalese; Nominalism; Objects, Properties, and Functions; Semantic Value.

Bibliography

- Albert of Saxony (1522). *Perutilis logica*. Venice, (Reprinted by Georg Olms, New York, 1974).
- Buridan J (2001). ‘*Summulae de Dialectica*’: an annotated translation with a philosophical introduction by Gyula Klima. New Haven, CT: Yale University Press.
- Carnap R (1959). ‘The elimination of metaphysics through logical analysis of language.’ In Ayer A J (ed.) *Logical positivism*. Glencoe, IL: Free Press. 60–81. (Original work published 1931.)
- Carnap R (1931). ‘Überwindung der Metaphysik durch logische Analyse der Sprache.’ *Erkenntnis* 2, 220–241.
- Bos E & Read S (2001). ‘Concepts: the treatises of Thomas of Cleves and Paul of Gelria.’ *Louvain-La-Neuve: Editions De L’institut Supérieur De Philosophie*, Louvain, Paris: Editions Peeters.
- Kretzmann N (1982). ‘Syncategoremata, exponibilia, sophistimata.’ In Kretzmann *et al.* (eds.). 211–245.
- Kretzmann N, Kenny A, Pinborg J & Stump E (eds.) (1982). *The Cambridge history of later medieval philosophy*. Cambridge, UK: Cambridge University Press.
- Locke J (1995). ‘An essay concerning human understanding,’ e-text, in public domain, <http://www.ilt.columbia.edu>, mark-up copyright © 1995 ILT Digital Classics. (Printed edition: *An essay concerning human understanding*, Nidditch P H ed., 1975.). Oxford: Oxford University Press.
- Peter of Ailly (1980). *Concepts and insolubles*. Spade P V, (trans.). Dordrecht, The Netherlands: Reidel.
- Peter of Spain (1992). *Syncategoremata*. First critical edition with an introduction and indexes by L M de Rijk, with an English translation by Joke Spruyt. Leiden: Brill.
- Pinborg J (1972). *Logik Und Semantik Im Mittelalter: Ein Überblick*. Stuttgart: Frommann-Holzboog.
- Sweeney E (2002, Winter). ‘Literary forms of medieval philosophy.’ In Zalta E N (ed.) *Stanford encyclopedia of philosophy*. Available at: <http://plato.stanford.edu>.
- William of Sherwood (1941). ‘Syncategoremata.’ O’Donnell J R (ed.). *Medieval Studies* 3, 46–93.
- Zupko J (2003, Spring). ‘Thomas of Erfurt.’ In Zalta E N (ed.) *The Stanford encyclopedia of philosophy*. Available at: <http://plato.stanford.edu>.

Syntax–Semantics Interface

J-P Koenig, University at Buffalo, Buffalo, NY, USA

© 2006 Elsevier Ltd. All rights reserved.

‘Syntax’ in this article refers to the organization of the words in a sentence and to the set of rules or constraints that organize words into sentences. ‘Semantics’ refers to the meaning of words and sentences and to the rules or constraints by which the meanings of sentences are built from the meanings of words. It is widely believed that the ability to combine words (and morphemes) to build more complex meanings from these syntactic combinations is a hallmark of human languages. Modeling this ability is one of the central aims in linguistics. The study of this ability is the study of the interface between syntax and semantics, and answers to this question distinguish the major current theoretical frameworks.

The Model of Perfection: Artificial Languages

To understand the issues that have occupied researchers, it is useful to compare natural languages with artificial or formal languages: Whether or how natural languages differ essentially from artificial languages helps classify current views. Artificial or formal languages are typically explicitly designed to be ‘perfect’ in the following sense. For each way of combining two or more syntactic expressions, there corresponds a unique way of combining the meaning of the expressions that are syntactically combined. Technically, an isomorphism can be defined between the syntactic and semantic components of formal languages. As a result, the meaning of formulas of arbitrary complexity is easily and unambiguously read from their syntactic surface organization. The formal language used to express referring expressions in arithmetic provides a simple example of this ‘perfection.’ The following are the syntactic formation rules for addition and multiplication.

- Syntactic rule 1: Number names and number variables are syntactically well-formed expressions called terms.
- Syntactic rule 2: If a and b are syntactically well-formed expressions, so is $\lceil a + b \rceil$.
- Syntactic rule 3: If a and b are syntactically well-formed expressions, so is $\lceil a * b \rceil$.

To each of these syntactic rules there corresponds a well-defined semantic rule.

- Semantic rule 1: Terms refer to numbers.
- Semantic rule 2: A formula of the form $(a + b)$ refers to the sum of the referent of a and the referent of b .
- Semantic rule 3: A formula of the form $(a * b)$ refers to the product of the referent of a and the referent of b .

The one-to-one correspondence between the syntactic rules and the semantic rules unambiguously indicates which computations are needed to obtain the referent of an expression of any complexity. Arithmetic expressions wear the needed computations to derive their referents on their sleeves, so to speak. As the two trees in Figure 1 informally depict, we can establish a one-to-one correspondence between syntactic computations and semantic computations. Such correspondences are known as homomorphisms or isomorphisms between syntax and semantics, depending on whether two different syntactic rules may or may not have the same semantic import. Only homomorphisms are relevant for natural languages.

The hypothesis that the interface between natural languages’ syntax and semantics works like that of formal languages goes back to the German mathematician Gottlob Frege (although it is not clear that Frege held that view; see Pelletier, 2001). Frege’s principle, the principle of compositionality, is defined as:

Principle 1. Compositionality Principle (Frege’s principle): The meaning of an expression is a function of the meaning of its parts and the way the parts are combined.

More operationally, since Montague (1974), the principle is understood to mean that we can establish a one-to-one correspondence or homomorphism between syntactic and semantic operations (see Janssen, 1997, for a good review). The models of the interface between the syntax and semantics of natural languages that have been proposed in the last 30 years can be profitably categorized on the basis of:

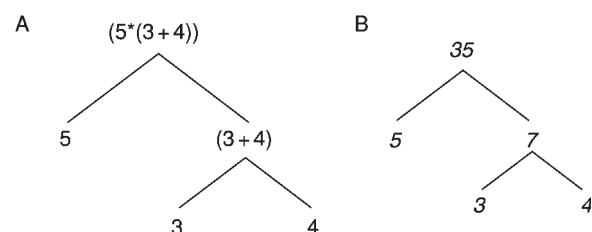


Figure 1 The parallel between (A) syntactic composition and (B) semantic composition in the language of arithmetic.

1. Whether they claim that natural languages differ **essentially** from the language of simple arithmetic formulas, that is, whether there is a homomorphism between the syntax and semantics of natural languages.
2. Whether the claimed correspondence (if any) must exist between syntactic **rules** and semantic **rules** (as Montague originally claimed) or between the syntactic and semantic **structures** that these rules output.

Where Natural Languages Seem Imperfect

Interfacing syntax and semantics would become a rather trivial task if natural languages' syntactic rules/constraints/structures paired with their semantic rules/constraints/structures as straightforwardly as in the case of the language of simple arithmetic. If natural languages are as 'perfect' as the language of simple arithmetic, they are not obviously so. Thus, claiming that there is no 'important theoretical difference' between formal and natural languages, as Montague (1974: 188) did, is a bold and noteworthy claim. Below are sentences that illustrate some of the mismatches (the 'imperfections') typical of natural languages. Discussions of how to appropriately model the interface between the syntax and semantics of these and similar classes of sentences has occupied much of the literature in the last 30 years.

(1a) It seems that John is tired.

(1b) John seems to be tired.

(2a) Who(m) did John see?

(2b) Jean a vu qui?
Jean has seen who(m)

(3a) Everyone loves someone.

(3b) $\exists x(\text{human}'(x) \wedge (\forall y(\text{human}'(y) \rightarrow \text{love}'(y, x)))$

(3c) $\forall y(\text{human}'(y) \rightarrow \exists x(\text{human}'(x) \wedge (\text{love}'(y, x)))$

(4a) Mary played.

(4b) Mary will play.

(5a) Mary began the beer.

(5b) Mary enjoyed the novel.

Sentence (1b) is an example of what is called (subject-to-subject) raising and, at least at a coarse level of analysis, is truth-conditionally equivalent to (1a). In both cases, the referent of *John* is the semantic argument of the meaning of the VP (*be*) *tired*; but only in sentence (1a) is *John* the subject of that VP. A consequence of this common semantics but distinct syntax is that the same combination of a subject and a VP does not seem to always have the same semantic import. In the case of the embedded clause of sentence

(1a), that syntactic combination means the application of the VP meaning to the subject's meaning; in the case of *John* and *seems to be tired* in sentence (1b), it appears that it does not because the referent of *John* does not appear to be a semantic argument of the meaning of *seems to be believed*. In other words, more than one semantic function might correspond to the combination of an NP and a VP.

The second example's relevance is slightly different. Let us assume that the meanings of questions expressed by sentences such as (2a) can be represented as in (6).

(6) for which animate x , John saw x

Then the surface structure of (2a) mirrors the semantic structure of (6). But the surface structure of the corresponding French sentence in (2b) (and other interrogative utterances in *in situ* languages) does not mirror its semantic structure (assuming that the semantics of English and French questions is the same). The difference is most easily seen in Figure 2. The syntactic position of *who* matches the semantic position of its translation in (6). Syntactically, *who* takes as its right sister a tree whose root is labeled S (its syntactic scope); semantically, the denotation of *who* is an operator that takes a propositional function as operand (its semantic scope). But *qui*, which is embedded within the verb phrase in French does not match its semantic scope (at least on the surface). No homomorphism seems to exist between syntactic and semantic structures for *in situ wh*-questions (although a homomorphism might still exist between syntactic and semantic rules).

Sentences such as (3a) have been at the center of many theories of the interface between the syntax and semantics of natural language. Such sentences are semantically ambiguous. According to one reading, there is at least one person that everybody loves, say *Gandhi*. Another reading is one in which, for each individual, there is at least one person he or she loves, but not all individuals need love the same entity; to each his or her own loved ones. The two first-order predicate logic formulas in (3b) and (3c) represent these two distinct readings. (*Everyone* and *someone* are understood here as equivalent to *every x that is*

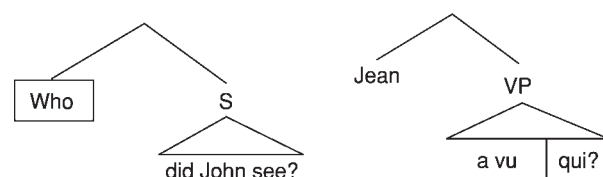


Figure 2 The difference between (A) English-style and (B) French-style questions.

human and *some x that is human*; *human'*, *love'* and *other'* expressions represent the meaning of the corresponding English words.) If we believe that these different readings correspond to different meanings of sentence (3a) (an assumption which may be challenged; see later discussion), sentences such as (3a) seem to contravene the claim that natural languages are ‘perfect’ because (3a) appears to have only one syntactic structure. We thus, have at least two semantic functions for combining a quantified direct object with the verb: one in which the semantic output is such that the subject has wide scope over it and one in which the semantic output does not force the subject to have wide scope over it. Contrary to the homomorphism hypotheses, there does not seem to be a unique semantic rule for every grammatical combination (here, the combination of verbs and direct objects).

The sentences in (4) exemplify the same problem as the sentences in (2) but within a single language. Under the assumption that, semantically, tense operators take propositions or event descriptions as operands (see Dowty, 1979; de Swart, 1998), the morpho-syntactic structure of (4a) does not match its semantic structure (even leaving aside the occurrence of the subject outside the verb phrase). The morpho-syntactic expression of the tense operator is a suffix on a word (the verb *play*), whereas its semantic translation takes a full event description or proposition as argument – something typically expressed as a VP or clause – as informally shown in (7).

(7) TENSE [EVENT DESCRIPTION/
PROPOSITION]

Example (4b) shows that English is not only ‘illogical,’ it is not systematic, in that the same semantic function (tense operators) can be expressed as an independent word that takes a verb phrase as complement or as a suffix on the verb.

Finally, example (5) illustrates that what is literally expressed does not provide all the intuitive semantic content we ascribe to utterances. What Mary began in (5a) is *drinking the beer* and in (5b) *reading the book* (possibly *writing the book*). Semantic content seems to have to be interpolated between the semantic translation of the verb and the semantic translation of the direct object, a phenomenon dubbed ‘coercion’ in the computational semantics literature (Moens and Steedman, 1988; Pustejovsky, 1995).

Theories of the Syntax–Semantics Mismatch

Existing grammatical frameworks make different hypotheses about how the syntactic–semantic mismatches

mentioned in the previous section can be resolved and about how artificial languages’ perfection is reflected in natural languages. We can distinguish at least three general hypotheses; each is currently actively being explored.

- Hypothesis 1. Deep Split Structural Isomorphism (DSSI): Natural languages’ imperfection is true of the surface syntactic structure of sentences. But strings can be assigned different structures at different levels of representation, and, for each kind of semantic relation, there is at least one level at which syntactic structures match very closely (if not entirely) the relevant aspects of semantic structures.
- Hypothesis 2. Natural Language Perfection (NLP): Natural languages’ imperfection is only apparent. A single semantic combination rule can in fact be assigned to every combination of syntactic expressions (although the syntactic category of natural language expressions might then need to be flexibly assigned).
- Hypothesis 3. Imperfections Reflect the Architecture of Grammars (IRAG): Natural languages’ imperfection is the result of their architecture (or even the architecture of the mind). Syntax and semantics are independent generative systems that are related through correspondence rules. Although there are regularities in how structures generated by one system are related to structures generated by the other, no homomorphic relation can be expected.

Proponents of these hypotheses typically focus on distinct subsets of the mismatches mentioned in the previous section. The DSSI hypothesis, for example, which goes back at least to the 1960s (see Lakoff, 1971), does not focus on coercion phenomena. The IRAG hypothesis, on the other hand, does not focus on quantifier scope mismatches. How easily each approach can tackle the **entire** set of mismatches or even whether they should (i.e., whether these mismatches form a natural class) is unclear at this point. This section focuses on the mismatches that each hypothesis models and, in particular, those mismatches whose analyses are easiest to present briefly.

The Deep Split Structural Isomorphism Hypothesis

The descriptive import of the DSSI hypothesis is most easily seen in typical analyses of examples (3a) and (4). Maintaining the DSSI hypothesis in the face of these surface mismatches leads to the claim that sentence (3a) has two distinct syntactic structures at the LF level of representation (often called L(ogical)-F(orm)), but a single syntactic structure at another level (often called s-structure) (see May,

1985; see Chomsky, 1995, and subsequent work in the Minimalist Program for a slightly different, but conceptually similar, model of these phenomena). Technically, these distinct structural analyses are generated by using two sets of rules, a set of context-free rules (e.g., those in (8) for sentence (3a)) and one or more additional (transformational) rules such as that in (9). In (8), the notation $\text{NP}[\pm Q]$ is used as a variable to range over all NPs, whereas $\text{NP}[+Q]$ ranges over only NP quantifiers. In (9), t_i is a placeholder that indicates where the NP quantifiers moved from and is called a trace. (These rules are adapted from Chierchia and McConnell-Ginet, 2000.)

- (8a) $S \rightarrow \text{NP}[\pm Q]\text{VP}$
 (8b) $\text{VP} \rightarrow \text{V NP}[\pm Q]\text{VP}$
 (8c) $\text{NP}[+Q] \rightarrow \text{Det N leveryone|someone}$
 (8d) $\text{Det} \rightarrow \text{every|a}$
 (8e) $\text{N} \rightarrow \text{man|woman}$
 (8f) $\text{V} \rightarrow \text{loves}$
 (9) $[_S \dots \text{NP}[+Q] \dots] \Rightarrow [_S \text{NP}[+Q]_i [_S \dots t_i \dots]]$

The context-free rules in (9) generate the single s-structure-level syntactic tree in Figure 3. Subsequent applications of the transformational rule in (9) results in the two LF-level syntactic trees in Figures 4 and 5, depending on which of the $\text{NP}[+Q]$ is targeted first. The two relevant readings of sentence (3a) now

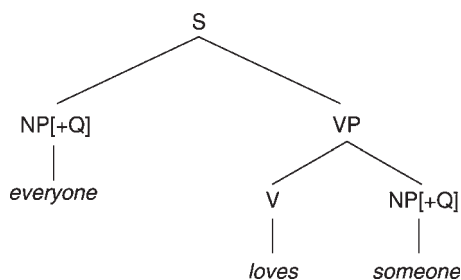


Figure 3 The s-structure representation of sentence (3a).

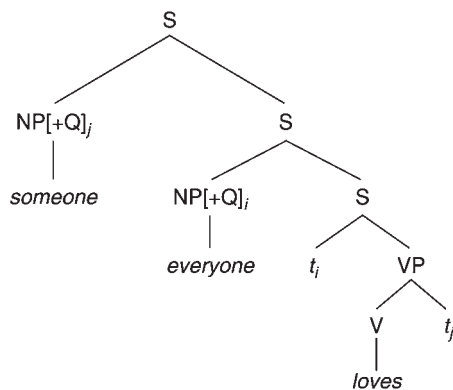


Figure 4 LF structure for the wide-scope existential reading of sentence (3a).

receive two distinct representations at one level of representation, and the homomorphism principle can be maintained if we assume the semantic interpretation of sentence (3a) is read from its LF representations. (Strictly speaking, homomorphisms require that ‘operations’ be defined in both domains put into correspondence, e.g., syntax and semantics. Thus, the output structures for which an homomorphism exists must be reanalyzed in terms of a local tree-building operation. Alternatively, the notion of order-homomorphism between structures outputted by the grammatical rules can be used, the relevant order being here tree-inclusion; see Cohn, 1981.) Compare the relative depth of embedding of the two quantified NPs in Figures 4 and 5 with that of the corresponding quantifiers in the first-order predicate logic formulas for the two readings of sentence (3a) in Figures 6 and 7.

The distinction in morpho-syntactic encoding that tense morphemes receive both within English and crosslinguistically is subject to a similar kind of analysis for proponents of the DSSI hypothesis. The homomorphism hypothesis can be maintained because a single surface expression can have two different structures at two distinct levels of representation. In particular, tense can have syntactic scope over the structure that denotes an event description (or

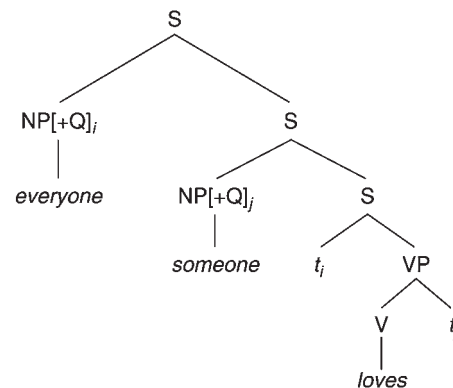


Figure 5 LF structure for the narrow-scope existential reading of sentence (3a).

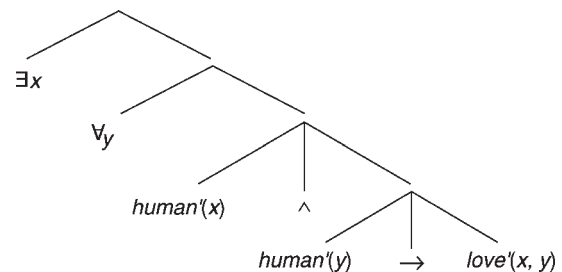


Figure 6 A tree representation of the wide-scope existential reading of sentence (3a).

proposition) at one level of representation (the d-structure, or its equivalent in the Minimalist Program) but not at another level of representation. The common structure of sentences (4a) and (4b) at the d-structure level is represented in Figure 8.

The distinct surface expression of the morpheme encoding tense in sentences (4a) and (4b) follows from operations similar to the transformational rule in (9) and the assumption that these operations are sensitive to the distinction between affixes such as *-ed* and free morphemes such as *will*. The contrast in the encoding of time in English is only the tip of the iceberg. Different tenses and various aspect markers may vary in their surface expression within a single language and across languages. The strategy just outlined can be applied generally to all these cases (see Cinque, 1999, for one of the most detailed explorations of this possibility).

Similar accounts are provided for the data set in (2). More generally, the claim that ‘deep down’ the interface between syntax and semantics is more transparent than what surface strings suggest requires providing each surface string with several distinct syntactic analyses:

- An analysis that allows predicate–argument relations (including predicate–argument relations of tense, aspect, and modal operators) to be homomorphically read from the syntactic tree (e.g., an analysis that reads the tense predicate/VP argument relation from the syntactic structures of trees such as the one in Figure 8, i.e., from structures before verb stems move).

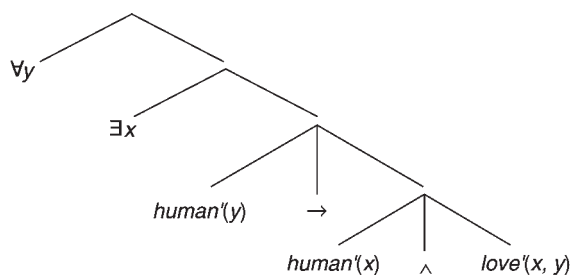


Figure 7 A tree representation of the narrow-scope existential reading of sentence (3a).

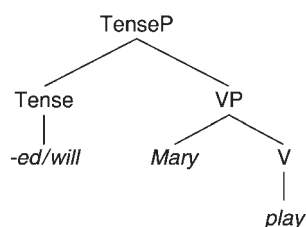


Figure 8 The d-structure representation of sentences in (4).

- An analysis that allows quantifiers and other scoping relations that do not involve a predicate–argument relation (focus vs. ‘presupposed,’ topic vs. comment, and so forth) to be homomorphically read from another syntactic tree (e.g., an analysis that reads the quantifier/scope relation from the syntactic structure of trees such as the ones in Figures 4 and 5 i.e., from structures that result from moving the NP quantifiers.)
- An analysis that represents the structure of the surface string.

To mediate among these analyses, successive applications of a general rule similar to (9) is posited. Morphemes, words, or phrases whose semantic interpretation contravenes their surface syntactic position must be moved to or from a position where the semantic interpretation of the local syntactic structure that they participate in can be easily read. For example, to account for the event description or propositional nature of the semantic argument of the past tense operator in sentence (4a), the past tense morpheme takes as syntactic complement at d-structure a structure that includes the verb stem for *play*, in conformity with its semantic constituency; subsequent movement of the verb stem ensures that it combines with the syntactically higher affix to form the surface verb form *played*. Similarly, the quantifiers *everyone* and/or *someone* in sentence (3a) are assumed to ‘covertly’ move to a position in which their semantic scope is easier to read from the local tree they are part of. As a result, the syntactic notion of hierarchical superiority (or c-command) matches the semantic notion of wide scope (at least in the traditional first-order predicate calculus representation of scope dependencies). Note that, the mapping of syntactic tree configurations onto semantic configurations must precede the movement operation in the case of the movement of the verb stem, but must follow the movement operation in the case of the movement of quantifiers. (Alternatively, the syntax-to-semantics mapping targets the position of the moved constituent, or head of a chain in the sense of Chomsky, 1981, for quantifiers, but targets the position of the trace of the moved constituent for verb stems and raised NPs such as *John* in (1b).)

In theories that follow the just-outlined general model of the interface between syntax and semantics, one syntactic structure can be provided from which the relevant part of semantic representations can be homomorphically read. But, there is no **single** such structure, a state of affairs that is seen by some as effectively abandoning the spirit of Frege’s compositionality principle (Jacobson, 2002). But, proponents of the DSSI hypothesis have stressed that the kind of

	loves	someone
	$(S \backslash NP) / NP :$	$(S \backslash NP) \backslash ((S \backslash NP) / NP) :$
	$\lambda y \lambda x.(\text{love}'(x, y))$	$\lambda q. \exists y(\text{human}'(y) \wedge q(y))$
Everyone	FA	
$S / (S \backslash NP) :$	$S \backslash NP :$	
$\lambda p. \forall x(\text{human}'(x) \rightarrow p(y))$	$\lambda x. \exists y(\text{human}'(y) \wedge (\text{love}'(x, y)))$	FA
$S :$		
$\forall x(\text{human}'(x) \rightarrow \exists y(\text{human}'(y) \wedge \text{love}'(x, y)))$		

Figure 9 Derivation of the semantic representation of the narrow-scope existential reading of sentence (3a).

structural operations and configurations necessary to model sentences such as (3a) or (4) are at play in syntactic processes that do not bear on the interface between syntax and semantics. Strictly syntactic constraints might thus provide independent evidence for the semantically more complex DSSI hypothesis (see May, 1985; Hornstein, 1995). Whether the adduced similarities between strictly syntactic constraints and the operations/structures presented in this section are sufficient to weaken Frege's principle remains a subject of debate.

The Natural Language Perfection Hypothesis

Hypothesis 2 answers the interface challenge by denying that natural languages have any real imperfections or that natural languages relate their syntax and semantics in a qualitatively distinct manner from artificial languages. The impression of imperfection comes merely from too simple a view of the rules available to combine natural language expressions and the erroneous reification of the notion of syntactic constituent structure. The compositionality of natural languages, in this view, pertains to the pairing of a set of semantic and syntactic rules (the rule-to-rule hypothesis of Bach, 1976), not to a pairing of the structures that these rules output. In contrast to the DDSI hypothesis, the NLP hypothesis claims that the homomorphism between syntax and semantics need not be split into two partial homomorphisms (to infer from surface expressions predicate–argument relations and operator–scope relations). It also claims that the syntax–semantics correspondence is best understood in terms of syntactic and semantic rules rather than between syntactic and semantic structures (local tree configurations, *modulo* the previously mentioned caveats). The NLP hypothesis goes back at least to Montague's work. It has been adopted by several different theories (e.g., Klein and Sag, 1985; Steedman, 2000; Copestake *et al.*, 2001); it is the hallmark of the Categorical Grammar/Lambek Calculus tradition. Steedman (2000) provides one of the most articulate versions of this approach. In his analysis of example (3a), as is common to all approaches

that assume that the translation from syntactic structure/categories to semantic structure/categories is local (the rule-to-rule hypothesis), each expression is assigned both a syntactic and semantic category, and each combinatorial rule affects both the syntactic and semantic categories of the expressions it combines. Syntactically, *loves* is taken to be a functor that must combine with a direct object NP to its right and a subject NP to its left to form a sentence. This syntactic subcategorization is written $(S \backslash NP) / NP$. For illustrative purposes, quantified NPs such as *everyone* and *someone* are treated syntactically as either NPs or functors that must combine with a VP to their right to form a sentence or with a verb to their left to form a VP. This second syntactic subcategorization is written $S / (S \backslash NP)$ and $(S \backslash NP) \backslash ((S \backslash NP) / NP)$, respectively. Semantically, *love* denotes a predicate missing two arguments, and *everyone* and *someone* denote predicates missing one argument each. Making use of λ -abstraction to model unsaturated argument positions, the meaning of *love* can be represented as:

$$\lambda y. \lambda x. \text{love}'(x, y)$$

and that of *everyone* and *someone* as

$$\lambda p. \forall x(\text{human}'(x) \rightarrow p(x))$$

and

$$\lambda q. \exists y(\text{human}'(y) \wedge p(y))$$

respectively. Three principles of combination are relevant to modeling sentences (3a) (a few more universal principles, plus language-specific constraints on these principles, are the only other things that are needed to model the interface between the syntax and semantics of natural languages, according to this view):

- Function application, the combination of an argument/complement and a predicate/functor: Combining (in that order) x and $y \backslash x$ or y / x and x yields a y .
- Type-raising, making an argument/complement into a predicate/functor: An expression of category x can be assigned to the category $y / (y \backslash x)$ or $y \backslash (y / x)$.

Everyone	loves	
$S/(S\backslash NP) :$	$(S\backslash NP)/NP :$	
$\lambda p.\forall x(human'(x) \rightarrow p(x))$	$\lambda y\lambda x.(love'(x, y))$	
FC		someone
$S/NP :$	NP	
$\lambda y.\forall x(human'(x) \rightarrow (love'(x, y)))$	$\lambda q.\exists y(human'(y) \wedge q(y))$	
FA		
$S :$		
$\exists y(human'(y) \wedge (\forall x(human'(x) \rightarrow love'(x, y))))$		

Figure 10 Derivation of the semantic representation of the wide-scope existential reading of sentence (3a).

- Function composition, composing two functors: Combining (in that order) two functors x/y and y/z yields x/z .

The derivations of the two relevant readings of sentence (3a) are given in **Figures 9** and **10**. (For ease of exposition, type-raising operations are assumed to have applied in the lexicon and are not represented.) The interpretation of sentence (3a) in which the existential quantifier has narrow scope over the universal quantifier follows from two uses of function application (FA in **Figure 9**; standard steps that simplify the resulting semantic formulas are omitted). The interpretation of sentence (3a) in which the existential quantifier has wide scope over the universal quantifier follows from the use of function composition first and then the use of functional application (FC and FA in **Figure 10**).

The two possible derivations of sentence (3a) outlined in **Figures 9** and **10** clearly obey Frege's principle and the rule-to-rule hypothesis. To each syntactic combination of two of its substrings corresponds a unique combination of the semantic representations of those two substrings. Leaving aside a few additional complexities, modeling the other mismatches previously mentioned, takes a similar form under this view. Alternative derivations (leading to either distinct surface expressions of the same semantic content or distinct semantic content for the same surface expressions) stem from the (at times nondeterministic) interaction of a handful of combinatory principles (function application, various forms of function composition, and type-raising). Proponents of the NLP hypothesis, thus, provide rule-based, truly compositional models of the interface between syntax and semantics. Whether operations posited by such models, like the operations/structures posited by DSSI models, have independent, purely syntactic justifications remains controversial.

The Imperfections Reflect the Architecture of Grammars Hypothesis

The IRAG hypothesis, which is more recent and less fleshed out than the previous two hypotheses, takes

the interface challenge to reflect a deep truth about the architecture of natural languages (if not the mind). In this view, natural languages are 'imperfect' because each level of analysis (syntax and semantics) has its own 'agenda,' so to speak, and although these 'agendas' are typically aligned, they need not be. The analysis of the sentences in (5) given in Jackendoff (1997) (see also Sadock, 1991, for an early proposal) best illustrates this hypothesis. The context-free rules in (8) account for the (simplified) syntactic structure of sentence (5a), as shown in the tree in **Figure 11**.

Two sets of correspondence rules link the syntactic structure diagrammed in **Figure 11** or that of the more explicit expression of its content in sentence (10a) and their common semantic representation. (*Begin* is assumed here to denote a two place predicate whose two arguments are an actor and a type of action; informally, $BEGIN'(x \text{ actor}, y \text{ action})$.)

(10a) Mary began drinking the beer

(10b) Mary began reading the book

The rules for interpreting the more explicit versions of sentences (5) in (10) are informally formulated as follows.

1. Interpret *begin* as $BEGIN'$.
2. Interpret the denotation of (definite) direct object NPs as semantic arguments of the denotation of verbs.
3. Interpret VP complements of verbs as denoting action arguments of the interpretation of verbs.

To model the 'nondefault' correspondence between the syntactic relation of the verb and its direct object and their semantic relation in sentences such as (5), Jackendoff posits rule 4 as an alternative to rule 2.

4. Interpret NP as $F(Int(NP))$ (i.e., as the unspecified activity denoted by F and involving the interpretation of NP).

The application of rules such as rule 4 must be further constrained so as to not apply in the 'default' case (see Sadock, 1990, for suggestions and Jackendoff, 2002, for more details).

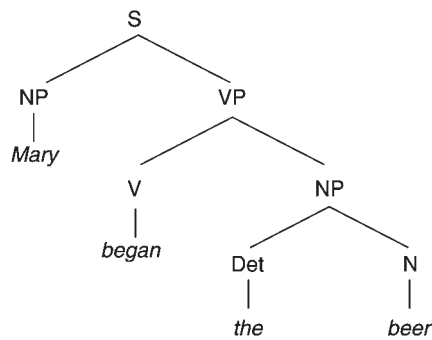


Figure 11 The syntactic structure of sentence (5a).

The IRAG hypothesis resembles the DSSI hypothesis in claiming that syntactic structures are linked to semantic structures (*modulo* the previously mentioned caveats) and that correspondence rules take the form: Configuration X in syntax corresponds to configuration Y in semantics. But it views the assumption that semantic structures can **always** be read from syntactic structures (as in the DSSI hypothesis) or the syntactic formation rules that have applied (as in the NLP hypothesis) as an illusion created by the existence of default correspondence rules. The IRAG hypothesis also resembles the NLP hypothesis in positing a direct pairing of syntactic and semantic types (NPs and $F(\text{Int}(\text{NP}))$ in rule 4, or the syntactic category and meaning of lexical items). But it insists that syntax and semantics are independent generative systems that derive their outputs on their own, and thus it views mismatches not as illusions but as evidence for this architectural independence or representational modularity (Jackendoff, 1997).

Shaking Things Up

How Specified Is Grammatical Meaning?

All the approaches discussed in the previous section presume that grammatical meaning (the output of interfacing the syntactic and the semantic components of a grammar) is a message whose content is similar to a logician's proposition (or its conceptual equivalent). The relevance of some of the examples in (2)–(5) to models of the interface between syntax and semantics depends on this presumption. Consider example (3a). The existence of one surface syntactic structure but two readings for this sentence is part of what motivates the DSSI hypothesis and the postulation of multiple syntactic levels of representations. Similarly, the two sequences of rules deriving the two readings of sentence (3a) sketched in Figures 9 and 10 presuppose that these two readings are the output of interfacing syntax and semantics. Examples of coercion such as (5) are only relevant to the uncovering of the architecture of grammar if the addition of the

'missing' material is believed to be a grammatical matter. If it is not, grammatical nondefault correspondence rules of the kind that rule 4 exemplifies are not needed to rectify the mismatch.

Recent work in computational semantics challenges traditional assumptions about the semantic output of grammars and, as a consequence, has altered the relevance of some of the examples in (2)–(5). A fair amount of work, in particular, challenges the view that the semantic output of grammar rules or constraints is fully specified and argues that this output is underspecified. (Cooper, 1983, is a precursor of that computational work. Philosophers of language and neo-Gricean semanticists have also challenged this view, e.g., Recanati, 1993; Levinson, 2001; some of their conclusions parallel those drawn by computational semanticists.) Two main areas of research have focused on whether scopal relations in sentences such as (3a) are grammatically underspecified and whether semantic type mismatches such as in (5) are resolved grammatically. The results of this research are:

- Many scopal relations may not be determined by grammatical rules or constraints, and their determination might involve postsemantic processes.
- The syntactic relation between a head and a complement (e.g., a verb and its direct object) may not correspond semantically to the relation between a predicate and an argument (even leaving aside quantificational NPs). The corresponding semantic relation may be indirect, as, for example, the relation between a predicate and an argument of an argument of ... an argument.

The difference between underspecified and fully specified semantic representations can be illustrated using tree diagrams. A fully specified semantic representation can be diagrammed as a set of fully connected trees, as in the case of Figures 6 and 7 for the first-order predicate calculus representations of the two relevant readings of sentence (3a). As discussed earlier, the DSSI hypothesis claims that the relative hierarchical position of the quantifiers in these semantic representations matches that of their syntactic expression at the LF level of analysis (see Figures 4 and 5). Similarly, sentence (11) (a slightly more complex version of (3a)) would receive as **one** of its LF analyses under the DSSI hypothesis the tree configuration in Figure 12.

(11) Every man loves a woman.

Underspecified semantic representations, on the other hand, can be diagrammed as a less than fully connected tree. The semantic output of parsing sentence (11), for example, would be a set of three partial trees, as Figure 13 illustrates.

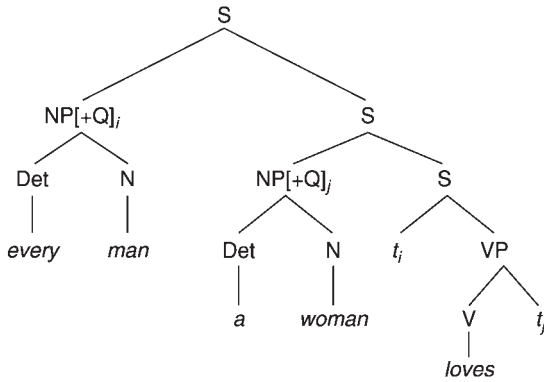


Figure 12 LF structure for the narrow-scope existential reading of sentence (11).

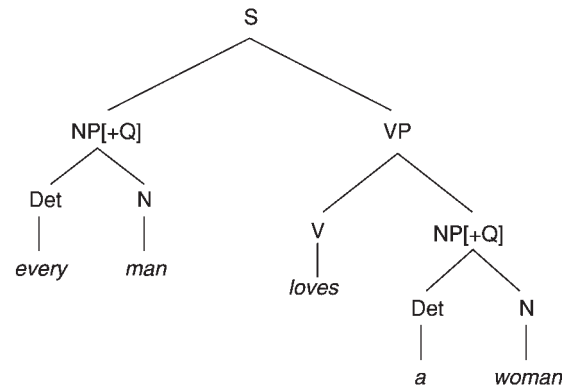


Figure 14 The syntactic structure of sentence (11).

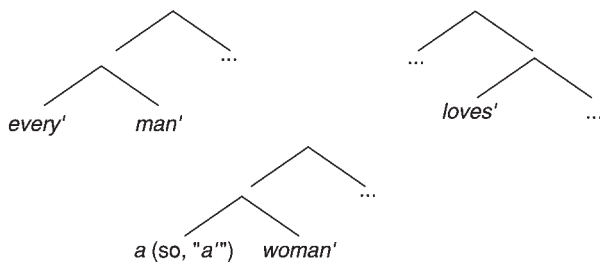


Figure 13 The underspecified semantics of sentence (11).

Each treelet in **Figure 13** can receive a compositional semantics (alternatively, each rule used to build the treelets can be paired with a semantic rule). The leftmost tree can be interpreted as:

$$\lambda p.(\forall x(\text{man}'(x) \rightarrow p(x)))$$

(the ... in the treelet is translated as a lambda-abstracted variable); the middle tree as:

$$\lambda q.(\exists y(\text{woman}' \wedge q(y)))$$

and the rightmost tree as:

$$\lambda y \lambda x.(\text{love}'(x, y))$$

This is very similar to what proponents of the NLP hypothesis would do: Each such treelet's meaning can be derived through functional application. But the semantic correlate of syntactically combining a verb and its direct object or a VP and its subject is merely the set union of these constituents' meanings. The semantic interpretation of these underspecified meanings is often said to be simply the set of fully connected semantic trees that can be built from these treelets. (Reyle, 1993, provides a possible direct model-theoretic interpretation of underspecified semantic representations.)

Now, because the semantic output of grammars is not a **single** fully connected tree, the structure of the

syntactic tree cannot reflect the structure of the semantic representation. Thus, such frameworks can assume a single syntactic structure as the output of the syntactic component of grammars – for example, the tree in **Figure 14** for sentence (11), given the phrase structure rules in (8) – while maintaining a rule-to-rule hypothesis.

The analysis of sentences that involve coercion follows parallel lines. Egg *et al.* (2001), for example, argued that the semantic interpretation of sentences such as (5) does not include additional semantic material (i.e., the direct object is not grammatically interpreted as *drinking the beer* and *reading the book*). Rather, the semantic output is a partial tree specification. The predicate *begin'* is required to dominate the interpretation of its complement (i.e., to include the interpretation of its complement within its second argument), but is not required to **immediately** dominate it, leaving open the possibility that pragmatic processes will interpolate conceptual content between the two. Nonstandard grammatical correspondence rules such as rule 4, whose sole purpose was to interpolate such conceptual content, would now become otiose.

Where Does the Meaning Come From?

Another assumption of the first two hypotheses we have reviewed is that the nature of the meaning carried by a lexical item differs from that encoded by syntactic structures or rules. Words (or morphemes) carry substantive meanings, whereas syntactic structures or rules only serve to combine or put these substantive meanings together. This division of labor is quite clear in the two derivations of sentence (3a) represented in **Figures 9** and **10**. The meanings of words carry the quantificational and predicative meanings; the rules only fill in the missing arguments in these lexical meanings (providing values for lambda-abstracted variables). Syntactic rules or

structures in such a view do not **carry** meaning; they merely glue or combine meanings carried by words or morphemes.

The last 20 years have seen the emergence of another view that holds that syntactic structures **do** carry meaning, too, at times (see Fillmore *et al.*, 1988; Ginzburg and Sag, 2001, for early and most articulated examples of this line of research). Syntactic structures thus can **both** glue meanings together and/or carry meaning on their own. Convincing evidence that syntactic structures, rules, or constructions carry meaning, is hard to come by because it is almost always possible to assign the meaning assigned to a syntactic construction to some lexical material within the sentence. But Ginzburg and Sag (2001) presented data (see also Huang, 1982) that suggest that syntactic structures do carry meaning. They show that fronting the *wh*-phrase, as English often does, as in (2a), has an effect on meaning. Fronted *wh*-phrases cannot have wider scope than their surface position. In contrast, *wh*-phrases that remain in position are freer in their scoping possibilities. In particular, they can have scope over the matrix clause when a question is embedded after verbs such as *know* and *ask*. The contrast between the Chinese (Mandarin Chinese) sentence in (12) (from Huang, 1982), in which the *wh*-phrase has remained in position, and its two possible interpretations indicated in its English translations illustrates the difference. The English *wh*-phrase has been fronted in both translations, and its scope is determined by its surface position. But the Chinese *wh*-phrase, which remains in place, can vary in scope. The different interpretative possibilities of the Chinese example and its English translations do not seem due to differences in **lexical** meaning between *shei* and its English counterpart *who*. Rather, it seems to follow from the different syntactic structures in which the two words occur (or the operations they undergo), thus suggesting that syntactic structures or rules contribute more to meaning than the glue for (substantive) lexical meanings.

- (12) [Zhangsan [zhidao shei mail-le shu]]
 Zhangsan know who buy-ASP book
 ‘who does Zhangsan know bought books? /
 Zhangsan knows who bought books’

Although much work has been devoted in the last 30 years to developing more precise architectures for the interface between syntax and semantics and finding data that support them, garnering convincing evidence that favors one hypothesis to the exclusion of the others has been hard. The difficulty partly stems from the fact that the proposed models are sufficiently rich to account for most of the data. Choosing among them is, therefore, more a matter

of elegance or some intuitive notion of simplicity than sheer descriptive adequacy. Furthermore, research has often focused on existence proofs (proof that some phenomenon can or cannot be accounted by one (version of) a model) or on relatively small grammar fragments for which explicit syntactic and semantic rules or constraints are provided. Two recent trends suggest better model evaluation might soon be available. First, several researchers now use data on human online sentence processing or word access as additional sources of evidence (see Steedman, 2000; Jackendoff, 2002). Second, large-scale implemented grammars that output both semantic and syntactic representations have developed in the last decade and might therefore allow the large-scale testing of the models on a realistic portion of natural languages, such as the ParGram project within Lexical-Functional Grammar and the LinGO project within Head-driven Phrase Structure Grammar.

Acknowledgments

I gratefully acknowledge Juergen Bohnemeyer, David Braun, Karin Michelson, Atsuko Nishiyama, and William Rapaport for their generous comments on a previous draft of this article. All remaining errors are mine.

Bibliography

- Bach E (1976). ‘An extension of classical transformational grammar.’ In *Proceedings of the 1976 Conference on Problems of Linguistic Metatheory*. Michigan State University Press. 183–224.
- Chierchia G & McConnell-Ginet S (2000). *Meaning and grammar* (2nd edn.). Cambridge, MA: MIT Press.
- Chomsky N (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky N (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Cinque G (1999). *Adverbs and functional heads: a cross-linguistic perspective*. New York: Oxford University Press.
- Cohn P (1981). *Universal algebra*. Dordrecht: Reidel.
- Cooper R (1983). *Quantification and syntactic theory*. Dordrecht: Reidel.
- Copestake A, Lascarides A & Flickinger D (2001). ‘An algebra for semantic construction in constraint-based grammars.’ In *Proceedings of the 39th ACL Meeting*. Toulouse, France.
- de Swart H (1998). ‘Aspect shift and coercion.’ *Natural Language and Linguistic Theory* 16(2), 347–385.
- Dowty D (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Egg M, Koller A & Niehren J (2001). ‘The constraint language for lambda structures.’ *Journal of Logic, Language, and Information* 10, 457–485.

- Fillmore C, Kay P & O'Connor C (1988). Regularity and idiomaticity in grammatical constructions: the case of *Let alone*. *Language* 64, 501–538.
- Ginzburg J & Sag I (2001). *Interrogative investigations: the form, meaning, and use of English interrogatives*. Stanford: CSLI Publications.
- Hornstein N (1995). *Logical form, from GB to minimalism*. Oxford: Blackwell.
- Huang J (1982). *Logical relations in Chinese and the theory of grammar*. Ph.D. diss., MIT.
- Jackendoff R (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff R (2002). *Foundations of language*. Oxford: Oxford University Press.
- Jacobson P (2002). 'The (dis)organization of the grammar: 25 years.' *Linguistics and Philosophy* 25, 601–626.
- Janssen T (1997). 'Compositionality.' In van Benthem J & ter Meulen A (eds.) *Handbook of logic and language*. Cambridge, MA: MIT Press. 417–473.
- Klein E & Sag I (1985). 'Type-driven translation.' *Linguistics and Philosophy* 8, 163–201.
- Lakoff G (1971). 'Generative semantics.' In Steinberg D & Jakobovits L (eds.) *Semantics: an interdisciplinary reader*. Cambridge, UK: Cambridge University Press. 232–296.
- Levinson S (2001). *Presumptive meanings*. Cambridge, MA: MIT Press.
- May R (1985). *Logical form: its structure and derivation*. Cambridge, MA: MIT Press.
- Moens M & Steedman M (1988). 'Temporal ontology and temporal reference.' *Computational Linguistics* 14(1), 15–28.
- Montague R (1974). *Formal philosophy*. New Haven, CT: Yale University Press.
- Pelletier F (2001). 'Did Frege believe Frege's principle.' *Journal of Logic, Language and Information* 10, 87–114.
- Pustejovsky J (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Recanati F (1993). *Direct reference: from language to thought*. Cambridge, UK: Blackwell.
- Reyle U (1993). 'Dealing with ambiguities by underspecification: construction, representation, and deduction.' *Journal of Semantics* 10, 123–179.
- Sadock J (1991). *Autolexical syntax*. Chicago: Chicago University Press.
- Steedman M (2000). *The syntactic process*. Cambridge, MA: MIT Press.

Relevant Websites

- <http://www2.parc.com> – ParGram project within Lexical-Functional Grammar.
- <http://lingo.stanford.edu> – LinGO project within Head-driven Phrase Structure Grammar.

Systematicity

P Robbins, Washington University in St. Louis,
St. Louis, MO, USA

© 2006 Elsevier Ltd. All rights reserved.

It is something of a commonplace that linguistic capacities are systematic: very roughly, that understanding a sentence entails understanding certain other sentences related to the first. Likewise, it is something of a commonplace that the best way to account for this fact is by supposing that linguistic meaning is compositional: again roughly, that to understand a sentence, it suffices that one understand the meanings of the words it contains and its syntactic structure. This explanatory hypothesis in turn accounts for much of the reason why compositionality is widely regarded by linguists and philosophers of language as a bedrock constraint on theories of meaning. But the real story about systematicity is rather more complicated. To begin with, the term 'systematicity' is far from univocal. A quick glance at the philosophical literature in which it has figured so prominently in recent years reveals at least three different senses of the term in common use. Since only confusion can result from running these different senses together, we will start by distinguishing them. Then

we will assess the plausibility of various systematicity claims and their broader theoretical significance, giving special attention to the question of compositionality.

Some Varieties of Systematicity

The first and arguably primary notion of systematicity concerns the capacity to process language. More specifically, our linguistic capacity is processing-systematic (p-systematic, for short) if and only if the capacity to understand (know the meaning of) a sentence s in a language L confers the capacity to understand (know the meaning of) a range of sentences $s_1^* \dots s_n^*$ in L such that each s_i^* is structurally related to s in an appropriate way (Fodor and Pylyshyn, 1988; Hadley, 1994; Cummins, 1996). For example, anyone who understands the sentence *John loves Mary* thereby also understands the sentences *Mary loves John*, *John loves John*, and *Mary loves Mary*. The latter three sentences are appropriately structurally related to the first sentence in virtue of the fact that each of the three sentences can be transformed into the first sentence by permuting or substituting co-occurring words of the same grammatical category

(here, the proper nouns *John* and *Mary*). A broader notion of p-systematicity can be arrived at by dropping the co-occurrence restriction, allowing for the substitution of co-categorical words not contained in the initial sentence. Structural relatives of *John loves Mary* in this more liberal sense include, in increasing order of distance, sentences like *John loves Elvis*, *Mary kicks Bridget*, and *Fred bombs Falluja*. The corresponding, stronger p-systematicity claim states that understanding a sentence s confers the capacity to understand a range of sentences $s_1^* \dots s_n^*$ such that each s_i^* is structurally related to s in the sense described above, and each word occurring in s_i^* is contained in the speaker-hearer's lexicon.

According to a second sense of the term, systematicity is a property of language itself, rather than a property of linguistic understanding. Since systematicity of this sort has to do with the expressive power of language, as opposed to the processing power of language users, we'll call it representation-systematicity. The basic idea is analogous to the notion of systematicity with which we began: a language L is representation-systematic (r-systematic) if and only if L can express a proposition p provided that L can express a range of propositions $p_1^* \dots p_n^*$ such that each p_i^* is structurally related to p in an appropriate way (Fodor and Lepore, 1991). The relevant structural relation on propositions can be illustrated with the same examples used above. Assuming L is r-systematic, if the proposition that *John loves Mary* is expressible in L , then so are the trio of propositions that *Mary loves John*, that *John loves John*, and that *Mary loves Mary*. Each member of the trio is structurally related to the initial proposition (that *John loves Mary*) because each can be transformed into the latter by permutation or substitution of co-occurring constituents expressed by words of the same grammatical category. As before, the class of propositions structurally related to the initial proposition can be expanded to include the propositions that *John loves Elvis*, that *Mary kicks Bridget*, and so on, by dropping the co-occurrence constraint on constituents. According to the strengthened r-systematicity thesis, a language can express a proposition p only if it can express a range of propositions $p_1^* \dots p_n^*$, where each p_i^* is structurally related to p in the manner described above, and each constituent of p_i^* is expressible by a word in the language.

Finally, a third notion of systematicity – what we will call grammar-systematicity – also denotes a property of language, but a different property than the one just described. Whereas representation-systematicity is a metasemantic property, grammar-systematicity has a broader metalinguistic scope, encompassing both syntax and semantics. A language L is grammar-systematic

(g-systematic) if and only if s is a grammatical sentence of L provided that any s^* structurally related to s is a grammatical sentence of L (Johnson, 2004). Since the relevant notion of structural relatedness between sentences is the same as that implicated in the definition of processing-systematicity, the same examples will serve. The fact that *John loves Mary* can be transformed without loss of grammaticality by permuting or substituting words of the same category, whether drawn from that sentence or from elsewhere in the lexicon, supports the claim that English is g-systematic.

Claims to the effect that natural language is systematic in any of these three ways – processing based, representation based, and grammar based – have a good deal of face plausibility. But before evaluating such claims, we need to scrutinize the above definitions more closely. In particular, we need to consider what it is for two words to belong to the same grammatical category. Pending some clarification of this issue, the concept of structural relatedness, and with it the very idea of systematicity, will remain somewhat obscure (Johnson, 2004). And a non-circular (i.e., systematicity independent) standard of sameness of category is not easy to formulate. Even the relatively fine-grained category of proper nouns, for example, seems too heterogeneous to support a robust claim of systematicity, as witnessed by the sentence *Julia sang the lead in 'West Side Story'* and its anomalous permutational variant *'West Side Story' sang the lead in Julia*. Given the plurality of constraints – syntactic, semantic, and phonological – on lexical substitution, better candidates for grammatical category-hood are going to be hard to come by (Harris, 1951).

How deep a problem this amounts to, however, depends to some extent on what notion of systematicity one is working with. In the case of grammar-systematicity, for example, the difficulty seems serious enough to warrant skepticism about the substance of the claim that language is g-systematic (Johnson, 2004). Processing-systematicity, on the other hand, is a different story. Since it is impossible to understand an anomalous sentence, it is plausible to suppose that understanding a sentence confers understanding of any structural variant of that sentence, provided that the variant in question is interpretable at all. Similar considerations apply to the case of representation-systematicity (Fodor and Lepore, 1991). It is plausible to suppose that the capacity of language to express a proposition entails the capacity to express any structural variant of that proposition, provided that the variant in question exists and is not anomalous. But this sort of strategy seems to be unavailable in the case of grammar-systematicity. Obviously, it won't do to say 'The structural variants of a grammatical sentence are grammatical, unless they are not.'

Be that as it may, it remains plausible that language is systematic in all three of the ways canvassed above, at least to a modest extent. Consider the following theses:

- *Minimal p-systematicity.* Understanding a sentence *s* entails understanding some structural variants of *s*.
- *Minimal r-systematicity.* If a language *L* can express a proposition *p*, then *L* can express some structural variants of *p*.
- *Minimal g-systematicity.* If *s* is a grammatical sentence, then some structural variants of *s* are grammatical.

Systematicity claims of this sort are hard to resist, given the apparent ease of generalization from the case of *John loves Mary* and its close variants. Whether language is p-systematic, r-systematic, or g-systematic in any stronger sense (e.g., whether ‘many’ can be strengthened to ‘all’) is an open question. But for the purposes of addressing the question of compositionality in natural language, these weak claims are enough to start with.

Explaining Systematicity

The idea that language is compositional, like the idea that language is systematic, can be formulated in various ways (see **Compositionality: Semantic Aspects**). According to one version, the principle of compositionality is a claim about our semantic competence, namely, that understanding the words in a sentence and the syntax of the sentence typically suffices for understanding the sentence. This is epistemic compositionality (e-compositionality). On another, more metaphysical, construal, the principle concerns language itself. It says that the meaning of a sentence typically supervenes on the meanings of the words in the sentence plus its syntax. We will call this ontic compositionality (o-compositionality). As noted above, systematicity is widely cited as a reason to believe in such principles. The connection between them, however, is more complex and attenuated than it may first appear. To start with, both e-compositionality and o-compositionality have to do mostly with semantics, and nothing else. As a result, considerations of g-systematicity, a broadly metalinguistic property, are of little significance. Instead, an argument to the best explanation from systematicity to compositionality will take one of two forms: either from p-systematicity to e-compositionality, or from r-systematicity to o-compositionality.

The line of argument from r-systematicity to o-compositionality is straightforward. We begin by assuming that a language *L* expresses the proposition *p*. This is equivalent to the claim that there is a sentence *s* in *L* that expresses *p*. According to

compositionality, the meaning of *s* is determined by the meanings of the words in *s* plus the syntax of the sentence. The same goes for structural variants of *s*, the meaning of which will be structural variants of *p*. If we suppose that *L* is compositional, then we get a ready explanation of the fact that *L* expresses a proposition *p* only if it expresses structural variants of *p*.

The line of argument from p-systematicity to e-compositionality is trickier. In order for the supposition that semantic competence is compositional to explain why that competence is systematic, it should follow from that supposition that the competence in question is systematic. But this is patently not the case. Assume that you understand a sentence *s*. The goal is to infer from this assumption that you understand structural variants of *s* as well. But the principle of e-compositionality says only that understanding the words and syntax of *s* suffices for understanding *s*; it does not say that understanding *s* suffices for understanding the words and syntax of *s*. Yet the latter assumption – the principle of reverse compositionality (Fodor and Lepore, 2001; Fodor, 2002) – is precisely what is needed in order to conclude, via the principle of compositionality, that understanding a sentence entails understanding its permutational variants. The amended argument is straightforward. If you understand *s*, then by reverse compositionality, you understand the syntax of *s* and the words occurring in *s*. These words can be re-ordered, *modulo* categorial constraints, within that syntactic structure to yield permutational variants of *s*. By compositionality, since you understand the words and syntactic structure of those variants, you also understand the variants. Hence, even if linguistic competence were compositional, that alone would not explain why it is systematic; one also needs to assume that it is reverse compositional.

A natural response to make at this juncture is that the phenomenon of p-systematicity warrants both e-compositionality and reverse e-compositionality. One problem with this move, however, is that the principle of reverse compositionality is probably false (Robbins, forthcoming). At a minimum, every theory of meaning on the market today, including the most austere versions of truth-conditional semantics, is incompatible with it (see **Meaning: Overview of Philosophical Theories**). For example, it’s possible to know the truth-conditions of the sentence *Pet fish have tails* without knowing that *pet* refers to pets and *fish* refers to fish. All one needs to know about the NP *pet fish* is that it refers to pet fish, and that fact is consistent with the assumption that *pet* refers to fish and *fish* refers to pets. Hence, if the truth-conditionalist story is correct, one can know the meaning of *Pet fish have tails* without knowing

the meanings of *pet* and *fish*. And non-truth-conditional accounts of meaning are in the same boat.

What's needed to escape this quandary, it seems, is a slight weakening of the thesis of minimal p-systematicity. Rather than saying that understanding a sentence strictly necessitates understanding some of its structural variants, we might say instead that instances of the first type of understanding are typically accompanied by instances of the second type. Such a pattern can be readily explained by the principle of compositionality together with a similarly watered-down form of reverse compositionality, according to which understanding a sentence is typically accompanied by an understanding of its syntax and the words contained therein. This in turn reveals how the systematicity of our semantic competence, however minimal in extent, militates in favor of its being compositional.

See also: Compositionality: Semantic Aspects; Meaning: Overview of Philosophical Theories.

Bibliography

- Cummins R (1996). 'Systematicity.' *Journal of Philosophy* 93, 591–614.
- Fodor J A (2002). 'The lexicon and the laundromat.' In Merlo P & Stevenson S (eds.) *The lexical basis of sentence processing*. Amsterdam: J. Benjamins.
- Fodor J A & Lepore E (1991). 'Why meaning (probably) isn't conceptual role.' *Mind and Language* 6, 328–343.
- Fodor J A & Lepore E (2001). 'Why compositionality won't go away: reflections on Horwich's 'deflationary' theory.' *Ratio* 14, 350–368.
- Fodor J A & Pylyshyn Z (1988). 'Connectionism and cognitive architecture: A critical analysis.' *Cognition* 28, 3–71.
- Hadley R F (1994). 'Systematicity revisited.' *Mind and Language* 9, 431–444.
- Harris Z (1951). *Methods in structural linguistics*. Chicago: University of Chicago Press.
- Johnson K (2004). 'On the systematicity of language and thought.' *Journal of Philosophy* 101, 111–139.
- Robbins P (forthcoming). 'The myth of reverse compositionality.' *Philosophical Studies*.

This page intentionally left blank

T

Tacit Knowledge

G S Rattan, University of Toronto, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

As Louise Antony and Norbert Hornstein (2003: 6) suggest, two of Chomsky's most significant philosophical contributions lie in his revival of mentalism in the philosophy of mind and rationalism in epistemology. Linguistics is the study of the system of knowledge possessed by speakers both innately (*see Innate Knowledge*) and in the relatively steady state condition of linguistic maturity. This knowledge is a kind of tacit knowledge. The aim of this article is to clarify some of the psychological and epistemic aspects of tacit knowledge, primarily as it is relevant to Chomskyan linguistics. As a psychological attitude, tacit knowing is distinguishable both from garden-variety kinds of propositional attitudes like believing and desiring (what I call full propositional attitudes) and from nonpropositional states or skills typically classified as instances of knowing how. The epistemic role of tacit knowledge in the broader issue of knowledge of meaning is discussed, and a more positive understanding of tacit knowing, with its distinctively epistemic gloss, is tentatively offered.

The Early Debate

It was clear to both Chomsky and his early critics that the kind of knowledge of language that speakers possess, if any, is not explicitly held. Speakers do not know explicitly the general principles to which human languages conform, or the grammars of the specific languages of which they are speakers. Chomsky and his early critics differed, however, on what they took to be the implications of that straightforward idea.

Critics argued that the invocation of psychological attitudes, especially knowing, was at best misleading and at worst wrong. The general objection is that the cognitive relation involved in the explanation of linguistic capacity, if indeed there is such a cognitive relation, cannot be that of knowing, because that cognitive relation does not sustain the connections

constitutive of knowledge (see, e.g., Stich, 1971: §4; Quine, 1972: 442; Devitt and Sterelny, 1999: 139; Dummett, 1991: 95–97). More specifically, early critics held that the states that underlie our ability to use language are, for most competent speakers, wholly unconscious, and when known, say by linguists, are known only indirectly through scientific theorizing. But, they contend, if one is to be ascribed a certain piece of knowledge, that knowledge should be recognizable 'from the inside' and not merely ascribable as a result of scientific theorizing. Let us say that having a full propositional attitude is constrained by a 'self-knowledge constraint,' according to which self-knowing is a distinctive, nontheoretical way of coming to know of one's attitudes and their contents.

Chomsky considers such objections (Chomsky, 1980, 1986) and replies as follows:

I have been speaking of "knowing English" as a mental state ... [But] to avoid terminological confusion, let me introduce a technical term devised for the purpose, namely "cognize," ... In fact I don't think that "cognize" is very far from "know" ... [Cognizing] is tacit or implicit knowledge, a concept that seems to me unobjectionable ... Cognizing has the structure and character of knowledge, but may be and in the interesting cases is inaccessible to consciousness. (Chomsky, 1980: 70–71)

The thrust of Chomsky's answer, then, is that if 'knowledge' offends, a technical term can be employed whose sense lacks precisely the offending features; but, continuing the response, the theoretical concept thereby introduced ought not to be thought of as fundamentally different from knowledge: it is knowledge, but it is unconscious, or not self-known.

Tacit Knowing vs. the Full Propositional Attitudes

We can see here an oversimplification of what might be involved in an account of the nature of propositional attitudes. There seems to be room to acknowledge both that speakers' knowledge of language does not sustain the connections constitutive of knowledge and the full propositional attitudes and

that, nevertheless, tacit knowledge is unobjectionable. The debate is oversimplified in turning only on a single, blunt constraint of self-knowledge. In what follows, I will try to enrich the debate by indicating considerations for thinking of tacit knowing as a propositional attitude without construing it as a full propositional attitude.

What more might be added to the understanding of the fully propositionally attitudinal to bring out its deep constitutive differences with tacit knowing? To begin with, we may take note of a pervasive feature of cognitive psychological theorizing, namely, that it construes the mind as modular: as segmented into mental components that use systems of encapsulated information that are inaccessible to conscious thinking, that are dedicated to representing highly structured or eccentric domains, and whose deployment is fast and mandatory (cf. Fodor, 1983). For the proponent of the idea that knowledge of language is a kind of tacit knowledge, the language faculty is itself, to a first approximation, such a modular system (Chomsky, 1986: chapter 1 [especially note 10]; Higginbotham, 1987; *see Cognitive Science and Philosophy of Language*).

The general properties of modular systems stand in sharp contrast with basic properties of the full propositional attitudes. The full propositional attitudes are inferentially integrated (Stich, 1978; Evans, 1981) and so can draw upon, and can be drawn upon, by a range of other full propositional attitudes (so they are neither encapsulated nor inaccessible) without regard to subject matter (so they are not dedicated). Although belief fixation may not be voluntary, it seems not to be mandatory in the sense in which the operation of modular systems is. The latter is a compulsion by psychological law; the former is something like a compulsion by reason (more on this below). Finally, the operation of the full propositional attitudes can be painfully slow. Reasoning can take a long time. These differences show that, on its own, the self-knowledge constraint is quite incomplete as to what is distinctive about the way that the information contained in the language faculty is held.

The point about modularity concerns primarily the nature of the attitude of tacitly knowing. But consider as well what Gareth Evans (1982) has called the 'generality constraint,' which pertains to the objects of the full propositional attitudes, thoughts or propositions, and their constituents, concepts (*see Concepts*). According to the generality constraint, concepts possess an inherent generality that mandates their recombining with appropriate concepts of other logical categories; slightly more formally, thoughts are closed under logico-syntactic formation

rules, up to conceptual incoherence or anomaly (cf. Peacocke, 1992: 42).

Now, no explanatory point seems to be served by imposing such a constraint on the representation of the information deployed in the language faculty. Speakers tacitly know the grammar that they do; knowledge of that grammar is deployed in some way or other, through the actions of mechanisms implementing algorithms that deploy the grammatical information tacitly known, so as to allow the acquisition, production, and perception of linguistic forms. Insisting that the constituents of the representation of linguistic information be subject to the generality constraint in no way illuminates the explanatory role of tacit knowing.

The reason can be clarified by thinking about the basic theoretical aims and motivations for both the full propositional attitudes and their contents. The full propositional attitudes and their contents are the fundamental theoretical entities in the conceptualization and explanation of the epistemic and practical successes and failures of agents. The explanations trace and assess the complexes of reasons for which agents believe and act, and implicitly evaluate those complexes, and agents, against an ideal of rational epistemic and practical functioning. The notions of epistemic and practical responsibility get their grip here, in the gap between actual and ideal. Since rational inference requires the interactions of attitudes and the recombination of concepts in a variety of ways, attitudes and contents must be such as to sustain the actual evaluations of epistemic status of agents in their inferential practices (Rattan, 2002: §4). The requirements that the full propositional attitudes be inferentially integrated and that concepts obey the generality constraint reflect these normative dimensions of the roles of attitudes and contents.

These ideas suggest, then, a deep point of contact with the self-knowledge constraint: that constraint will be relevant, like inferential integration and generality, insofar as it reflects fundamental features of rational practice. And surely it does. Reflecting on one's attitudes is a way of increasing the rational status of those attitudes. But if reflection plays that rational role, then it must be that our access to our attitudes is reliable and possessed of entitlement (cf. Burge, 1996); our access to our own minds must in the most basic cases be knowledge. Demanding that the attitudes and their contents be self-known, then, is part of a general account of the full propositional attitudes and their contents that construes them so that they may play their role in conceptualizing and explaining the normative statuses of agents.

Tacit Knowing vs. Knowing How

But it may be objected that for all that has been said so far, tacit knowing may not be a kind of content-bearing state at all; given the deep differences between tacit knowing and the full propositional attitudes, tacit knowing simply seems not to be a genuine mental attitude toward contents.

The objection can be elaborated by insisting that speaking a language is not a matter of knowing propositions at all. As Michael Devitt and Kim Sterelny put it:

[C]ompetence in a language does not consist in the speaker's semantic propositional knowledge of or representation of rules. It is a set of skills or abilities ... It consists in the speaker being able to do things with a language, not in his having thoughts about it. (Devitt and Sterelny, 1999: 187)

Perhaps there are no states of knowing the facts about the language – no knowing the rules or grammar of the language. A speaker no more knows the rules of the grammar than one who can ride a bike knows the laws of mechanics governing balance and maneuver. The language faculty is a faculty of knowing how to speak a language. Call the proponent of such a view the 'knowing-how theorist.'

Of course it cannot be denied that speakers do know how to speak the language; but the knowing-how theorist must justify the idea that attributions of knowing **how** are not just species of attributions of knowing **that**, as certain syntactic and semantic evidence suggests (Stanley and Williamson, 2001). Again, it looks as though the knowing-how theorist simply misses the point: although speakers know how to speak the language, that is not an explanation, but a description of what needs to be explained (cf. Higginbotham, 1994). But let us suppose that the knowing-how theorist is proposing an alternative explanation, one that does not appeal to tacit knowing. The knowing-how theorist has at least two significant hurdles to overcome.

First, suppose it is granted that tacit knowing does not have the direct role in rationalizing conceptualizations and explanations that the full propositional attitudes do. There still seems to be nothing to prevent thinking of tacit knowings as having the same attitude-content structure that the full propositional attitudes do. The attitude is not, admittedly, understood as playing a certain rational role, but it still may have a regular and lawlike causal role, and that would seem to be enough to think of there being something attitude-like – a distinctive way in which the information is held – in the picture. (Indeed this is the usual

understanding of the attitudes in causal functionalism; see, for example, Lewis, 1972.) This, for example, is what we might like to say about animals and their attitudes. They fail to have states that satisfy the normative constraints that the full propositional attitudes do, but it would be theoretically heavy handed to say that they do not have states that are very much belief- and desire-like.

Again, serving the aims of rationalizing explanations may require that contents be individuated at the level of Fregean sense rather than reference; this is the point of 'Frege's Puzzle' (Frege, 1892). But if tacit knowings fail to figure directly in these kinds of rational phenomena, then their contents may legitimately be exempted from having the general properties that the contents of attitudes that do so figure must have. The contents of tacit knowings may be Russellian, or have a limited need for cognitive difference without difference in reference (see **Sense and Reference: Philosophical Aspects**). These constituents of contents, objects and properties, presumably do not obey the generality constraint, so there are ways in which contents may be involved yet fail to be like the contents of the full propositional attitudes. The issues here are complex, and I mean only to flag the general issue about the individuation of the contents of tacit knowings and how that may serve to distinguish them from the contents of the full propositional attitudes.

Second, perhaps the proponent of knowing how thinks that the explanation of linguistic ability is merely dispositional, like the explanation of the shattering of a glass by appeal to its fragility. Here, there is some categorical, microstructural, property of the glass that, simplifying tremendously, explains why in conditions of the appropriate sort, the glass shatters. In a like manner the knowing-how theorist may appeal to the categorical neurophysiological ground of linguistic dispositions as comprising a nonintentional explanatory level. This idea, however, is subject to all the general objections that favor functional explanations over neurophysiological ones. It neglects a tradition of thinking about psychological explanations as involving multiple – computational, informational, algorithmic, and implementational – levels (see Marr, 1982; Peacocke, 1986). Indeed the functional states will be realized by neurophysiological states, but the explanation will be cashed out at some level that abstracts from neurophysiological description.

What Is Tacit Knowledge?

So far tacit knowing has been negatively characterized, by being distinguished from both the full

propositional attitudes and knowing how. But how are we to understand the nature of tacit knowledge in more positive terms? Tacit knowing plays a role in the explanations of linguistic acquisition, perception, and production. The described capacities are epistemic in character. Their explanations are part of more elaborate explanations that seek to understand the general phenomenon of communication, including its substantial pragmatic elements. To keep the discussion manageable, I will ignore the more elaborate problems. As well, I will focus only on the problems of linguistic perception and linguistic production. I mean only to gesture roughly at the kinds of considerations that are involved.

According to these explanations, speakers are able to make knowledgeable sound-meaning pairings for sentences because they have information about the phonological, syntactic, and semantic – grammatical – properties of the expression-types that make up those sentences. This information is drawn on by perceptual and production mechanisms. Simplifying immensely, production mechanisms take inputs from intentions to say that *p*, and make available to a speaker, through the use of the grammatical information, expression types that mean that *p*. Perceptual mechanisms take acoustic inputs and, through the use of grammatical information, impose grammatical properties on them, eventuating in an experience of meaning, on the basis of which meaning judgments are made (cf. Fricker, 2003). A fully fleshed out epistemology of meaning would explain the epistemic differences between knowing one's own meaning in production and knowing others' meanings in perception by, for example, considering how the inputs to production and perception, respectively, allow for different kinds of mistakes in the eventuating judgments about meaning.

Let us assume that information is deployed in speakers' linguistic epistemic achievements (so these achievements are not examples of knowing how), and that this information is tacitly rather than explicitly held (so it is not an example of a full propositional attitude). Still, why must we accept that this information is known rather than just truly believed? The question is difficult and fundamental. I offer here some potential lines for understanding.

Suppose, as I have been suggesting, that tacitly held information or content is involved in the explanation of linguistic capacities. Two things are of note here about this information. First, these representations have been formed by a reliable mechanism – one that uses speakers' innate representations of Universal Grammar – that reproduces the grammatical information represented in the minds of speakers in one's community. In normal environments, acquiring

these representations will equip one to come to judge knowledgeably the meanings of other speakers. Second, once one moves away from folk conceptions of public language, it is plausible to think of the facts about which language one speaks as settled by one's grammatical representations or I-language (Chomsky, 1986: Chapter 2; Higginbotham, 1991; Barber, 2001; *see E-Language versus I-Language*). Judgments about what one means oneself will then be reliably produced, again as outlined above; and since the facts about one's language are determined by one's grammatical representations, they will be reliably produced by the facts that determine the language one speaks. It seems that as a phenomenon at the level of the full propositional attitudes, knowing our own meanings and knowing the meanings of others, when we do, is not an accident.

We can think of the foregoing as giving the outlines of a philosophical explanation of what might be called the success-presupposing features of the explanation of our linguistic capacities. The explanations are not success neutral (see Burge, 1986, who attributes the phrase to Bernard Kobes), in that the explanations are explanations of epistemic capacities that are generally presumed to be successful. But if that is right, we are in a position to say something about why the information tacitly held is knowledge. Unless that information were known, rather than just truly believed, it would seem to be a mystery why drawing on that information in perception and production leads in general to knowledge. So one suggestion is that the status of the information as knowledge comes from the distinctive explanatory role of the tacitly held information in explanations of our generally epistemically successful linguistic capacity.

That's one pass at vindicating the attribution of knowledge. But perhaps something deeper can be said. Here we can return to Chomsky's rationalism. Sometimes Chomsky's rationalism seems to be a genetic rationalism that emphasizes the innate character of Universal Grammar, vindicating an early modern doctrine of innate ideas (*see Innate Ideas*). But there is another way to think of this rationalism, in which it emerges as a more full-blooded epistemic doctrine. In this way of thinking, at some point in the evolution of humankind, minds came into a cognitive relation with certain abstract structures, with very valuable combinatorial properties. These abstract structures are languages. We have already seen that the tacit is not the realm of epistemic and practical responsibility. So the status of knowledge for tacit representations will not accrue as a result of some personal-level achievement. The status of knowledge for the representations that underlie our linguistic capacity derives instead from a

natural attunement of the modular structures of the human mind to the abstract combinatorial structures of language.

See also: Cognitive Science and Philosophy of Language; Concepts; E-Language versus I-Language; Innate Ideas; Innate Knowledge; Sense and Reference: Philosophical Aspects.

Bibliography

- Antony L & Hornstein N (2003). *Chomsky and his critics*. Oxford: Blackwell.
- Barber A (2001). 'Idiolectal error.' *Mind and Language* 16, 263–283.
- Burge T (1986). 'Individualism and psychology.' *Philosophical Review* 95, 3–45.
- Burge T (1996). 'Our entitlement to self-knowledge.' *Proceedings of the Aristotelian Society* 96, 91–116.
- Chomsky N (1980). *Rules and representations*. New York: Columbia University Press.
- Chomsky N (1986). *Knowledge of language: its nature, origin, and use*. Westport CT: Praeger.
- Chomsky N (1991). 'Linguistics and cognitive science: problems and mysteries.' In Kasher A (ed.) *The Chomskyan turn*. Oxford: Blackwell. 3–25.
- Devitt M & Sterelny K (1999). *Language and reality: an introduction to the philosophy of language* (2nd edn.). Oxford: Blackwell/MIT Press.
- Dummett M (1991). *The logical basis of metaphysics*. Cambridge MA: Harvard University Press.
- Evans G (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Evans G (1985/[1981]). 'Semantic theory and tacit knowledge.' Reprinted in his *Collected Papers*, 322–342.

- Fodor J (1983). *The modularity of mind*. Cambridge MA: MIT Press.
- Frege G (1892). 'On sense and reference.' In Beaney M (ed.) *The Frege reader*. Malden MA: Blackwell. 151–180.
- Fricker E (2003). 'Understanding and knowledge of what is said.' In Barber A (ed.) *Epistemology of language*. Oxford: Oxford University Press.
- Higginbotham J (1987). 'The autonomy of syntax and semantics.' In Garfield J (ed.) *Modularity in knowledge representation and natural language understanding*. Cambridge, MA: MIT Press.
- Higginbotham J (1991). 'Remarks on the metaphysics of linguistics.' *Linguistics and Philosophy* 14, 555–566.
- Higginbotham J (1994). 'Priorities in the philosophy of thought.' *Supplementary Proceedings of the Aristotelian Society* 78, 85–106.
- Lewis D (1972). 'Psychophysical and theoretical identifications.' *Australasian Journal of Philosophy* 50, 249–258.
- Marr D (1982). *Vision*. New York: W. H. Freeman & Company.
- Peacocke C (1986). 'Explanation in computational psychology: language perception and level 1.5.' *Mind and Language* 1, 101–123.
- Peacocke C (1992). *A study of concepts*. Cambridge, MA: MIT Press.
- Quine W V O (1970). 'Methodological reflections on current linguistic theory.' *Synthese* 21, 386–398.
- Rattan G S (2002). 'Tacit knowledge of grammar: a reply to Knowles.' *Philosophical Psychology* 15, 135–154.
- Stanley J & Williamson T (2001). 'Knowing how.' *Journal of Philosophy* 98, 411–444.
- Stich S (1971). 'What every speaker knows.' *Philosophical Review* 80(4), 476–496.
- Stich S (1978). 'Beliefs and subdoxastic states.' *Philosophy of Science* 45, 499–518.

Temporal Logic

A Galton, University of Exeter, Exeter, UK

© 2006 Elsevier Ltd. All rights reserved.

Tense Logic

Tense Logic was introduced by Arthur Prior (1957, 1967) as an aid to elucidating various philosophical problems concerning time. Because it was used to provide formal renderings of propositions originally formulated as natural language sentences, it was natural to consider Tense Logic also as a tool for linguistic analysis in the context of linguistics rather than philosophy. As such, it was soon found to have considerable limitations, which stimulated the

development of extensions or alternatives to Tense Logic as it had been conceived originally by Prior.

Syntax of Priorean Tense Logic

Prior took the basic expressions of temporality in ordinary language to be the past, present, and future tenses. For the past and the future, he introduced proposition-forming operators *P* and *F*, called **tense operators**, which, when prefixed to any proposition ϕ , yielded propositions that expressed, respectively, the past-tense and future-tense versions of ϕ , for example:

- ϕ : It is raining.
- $P\phi$: It was raining.
- $F\phi$: It will be raining.

(More often, in fact, Prior would render $P\phi$ in the form ‘It has been raining’, using the present perfect rather than the simple past. In English, this reads more naturally because ‘It was raining’ seems to refer implicitly to some particular time assumed known.)

Used in combination with the standard connectives of Propositional Logic ($\wedge, \vee, \rightarrow, \leftrightarrow, \neg$), Prior’s tense operators allowed one to express temporal relationships, e.g., using ϕ for ‘John arrives’ and ψ for ‘Mary leaves’,

- $P(\phi \wedge \psi)$: John arrived when Mary left.
- $P(\phi \wedge P\psi)$: When John arrived, Mary had left.
- $F(\phi \wedge P\psi)$: When John arrives, Mary will have left.

This kind of analysis is too crude to capture the nuances of expression possible in natural language. Whereas the formulae $P(\phi \wedge \psi)$ and $P(\psi \wedge \phi)$ are logically equivalent, the sentences ‘John arrived when Mary left’ and ‘Mary left when John arrived’ are not. It might be argued, however, that the two sentences share some ‘core’ semantic content, and that this is what is expressed by the logical formulae. On this basis, Tense Logic is worth studying for the clarity it affords in analyzing at least some aspects of the linguistic expression of time.

Of particular interest is the interaction of the tense operators with negation. The formula $\neg P\phi$ says it has never been true that ϕ ; it is not equivalent to $P\neg\phi$, which merely says it has (at least once) been false that ϕ . Reading ϕ as ‘The prime minister of the UK is a woman’, $\neg P\phi$ is false (because of Margaret Thatcher), whereas $P\neg\phi$ is true (because of all the other prime ministers). The negation of the latter, $\neg P\neg\phi$, says it has never been false that ϕ , and hence that ϕ has always been true. Prior introduced a new operator, H , as an abbreviation for the sequence $\neg P\neg$, and analogously G for $\neg F\neg$. The full set of operators is now

- $P\phi$: It has been the case that ϕ (on at least one occasion).
- $H\phi$: It has always been the case that ϕ .
- $F\phi$: It will be the case that ϕ (on at least one occasion).
- $G\phi$: It will always be the case that ϕ .

Using these, Prior could enunciate tense-logical theses such as $\phi \rightarrow GP\phi$, which says that if ϕ is true now then it will always be the case that ϕ has been true.

Semantics of Tense Logic

To express the meaning of the tense operators, we assume that propositions are evaluated as true and false at times, that a proposition may be true at some

times and false at others, and that times are ordered by the relations *earlier than* and *later than*:

- $P\phi$ is true at time t if and only if ϕ is true at some time t' earlier than t .
- $H\phi$ is true at time t if and only if ϕ is true at every time t' earlier than t .
- $F\phi$ is true at time t if and only if ϕ is true at some time t' later than t .
- $G\phi$ is true at time t if and only if ϕ is true at every time t' later than t .

In the standard formal semantics for Tense Logic, a temporal frame is a pair $(T, <)$, where T is a set of times, and $<$ is a binary relation on T . For $t, t' \in T$, we understand $t < t'$ as the formal counterpart of the relation ‘ t is earlier than t' ’. An interpretation \mathcal{I} over frame $\mathcal{F} = (T, <)$ assigns a truth value to each atomic formula at each time in the frame. The fundamental semantic relation is satisfaction of formula ϕ at time t in frame \mathcal{F} under interpretation \mathcal{I} , symbolized as $\mathcal{F}, t \models_{\mathcal{I}} \phi$. If this does not hold, we write $\mathcal{F}, t \not\models_{\mathcal{I}} \phi$. We assume that all such relations are known for each *atomic* formula at each time; the truth values of compound formulae are determined by the semantic rules:

- $\mathcal{F}, t \models_{\mathcal{I}} P\phi$ if and only if $\mathcal{F}, t' \models_{\mathcal{I}} \phi$ for some $t' \in T$ such that $t' < t$.
- $\mathcal{F}, t \models_{\mathcal{I}} H\phi$ if and only if $\mathcal{F}, t' \models_{\mathcal{I}} \phi$ for every $t' \in T$ such that $t' < t$.
- $\mathcal{F}, t \models_{\mathcal{I}} F\phi$ if and only if $\mathcal{F}, t' \models_{\mathcal{I}} \phi$ for some $t' \in T$ such that $t < t'$.
- $\mathcal{F}, t \models_{\mathcal{I}} G\phi$ if and only if $\mathcal{F}, t' \models_{\mathcal{I}} \phi$ for every $t' \in T$ such that $t < t'$.

(For Boolean compounds with \wedge, \vee , etc., the standard rules are used.)

Some tense-logical formulae can now be proved to be universally valid, i.e., satisfied at every time in every frame. An example is Prior’s formula $\phi \rightarrow GP\phi$. For suppose $\mathcal{F}, t \models_{\mathcal{I}} \phi$, and let $t' \in T$ be any time such that $t < t'$. Then the semantic rule for P implies $\mathcal{F}, t' \models_{\mathcal{I}} P\phi$. Because this holds for every $t' \in T$ such that $t < t'$, the rule for G gives us $\mathcal{F}, t \models_{\mathcal{I}} GP\phi$, as required. The formulae provable in this way constitute a system known as Minimal Tense Logic, generally designated K_t .

Because the class of all temporal frames imposes no constraints on the nature of the $<$ relation, K_t does not provide an adequate basis for reasoning about temporal relationships. It is therefore usual to restrict the class of frames by requiring $<$ to satisfy certain structural properties such as transitivity, density, or linearity. A transitive frame $(T, <)$, for example, obeys the rule

$$\text{TRANS } \forall t, t', t'' \in T (t < t' \wedge t' < t'' \rightarrow t < t'')$$

i.e., if t is earlier than t' , which in turn is earlier than t'' , then t must be earlier than t'' . It can be shown that TRANS characterizes transitive frames in the following sense:

The frame $\mathcal{F} = (T, \prec)$ is transitive if and only if
 $\mathcal{F}, t \models_{\mathcal{I}} \text{TRANS}$ for every interpretation \mathcal{I} over \mathcal{F} ,
 and every time $t \in T$.

Much attention has been given to the problem of finding tense-logical formulae which correspond to first-order frame properties just as TRANS corresponds to transitivity (Van Benthem, 1991). Not all frame properties have corresponding tense-logical formulae, and conversely, not all tense-logical formulae correspond to first-order frame properties. Thus, there is no first-order characterization of the frames for which $GF\phi \rightarrow FG\phi$ holds; and no tense-logical formula exactly characterizes the *irreflexive* frames, which satisfy $\forall t \neg(t \prec t)$.

Extensions of Tense Logic

Increasing the Expressive Power: 'Since' and 'Until'

Sentences such as *John has been happy since Mary arrived* and *John will be happy until Mary arrives* cannot be expressed using Prior's operators. In 1968, Hans Kamp introduced binary operators S and U with the following semantics:

$\mathcal{F}, t \models_{\mathcal{I}} S\phi\psi$ if and only if there is a time
 $t' \prec t$ such that
 (i) $\mathcal{F}, t' \models \phi$, and
 (ii) $\mathcal{F}, t'' \models \psi$ for every $t'' \in T$ such that $t' \prec t'' \prec t$.
 $\mathcal{F}, t \models_{\mathcal{I}} U\phi\psi$ if and only if there is a time
 $t' \prec t$ such that
 (i) $\mathcal{F}, t' \models \phi$, and
 (ii) $\mathcal{F}, t'' \models \psi$ for every $t'' \in T$ such that $t \prec t'' \prec t'$.

Let ϕ stand for 'Mary arrives' and ψ for 'John is happy'. Then $S\phi\psi$ is true now so long as Mary arrived at some past time, and John was happy at every time between then and now – in other words, John has been happy since Mary arrived. Similarly, $U\phi\psi$ is true now if Mary will arrive at some future time, and John will be happy at every time between now and then. This implies John will be happy until Mary arrives, but is not equivalent to it, because the English sentence would also be true if Mary never arrives, so long as John remains happy forever. This can be expressed as $U\phi\psi \vee G\psi$.

Kamp showed that for frames in which the flow of time is ordered like either real numbers or integers, the operators S and U suffice to express every first-order property expressible by means of \prec and quantification over times. In this sense, the logic of S and

U is *expressively complete* in a way that the logic of P , F , G , and H is not.

The Indeterminate Future

Although it is natural to suppose that a temporal frame should be linear, the need has frequently been felt for temporal frames with a branching structure. In a linear frame, given two distinct times, one must precede the other: $\forall t, t' \in T (t = t' \vee t \prec t' \vee t' \prec t)$. The most popular form of non-linear structure is a *future-branching* frame, in which each time has a unique linear past, but may have many distinct futures. This is secured by the 'left-linearity' axiom that restricts linearity to times in the past of a given time: $\forall t, t', t'' (t \prec t'' \wedge t' \prec t'' \rightarrow t = t' \vee t \prec t' \vee t' \prec t)$.

This kind of structure supports reasoning about alternative *possible futures*: it captures the idea that whereas the past is fixed, the future is indeterminate. This means that some future events are inevitable (occurring in every possible future) whereas others are merely possible (occurring in some but not all possible futures). To quantify possible futures, we introduce modal operators. One way of doing this is as follows: $\Diamond\phi$ says there is a possible future which makes ϕ true now (so ϕ is possible), while $\Box\phi$ says that every possible future makes ϕ true now (so ϕ is necessary). The semantics makes explicit reference to possible futures; a *history* is any maximal linear subset of T , and formulae are evaluated with respect to pairs (H, t) , where H is a history and $t \in H$. Semantic rules for the existing operators can be straightforwardly rewritten by incorporating a single history term, as e.g.,

$\mathcal{F}, H, t \models_{\mathcal{I}} F\phi$ if and only if $\mathcal{F}, H, t' \models_{\mathcal{I}} \phi$ for some
 $t' \in H$ such that $t \prec t'$.

For the new modal operators, we have the rules

$\mathcal{F}, H, t \models_{\mathcal{I}} \Diamond\phi$ if and only if $\mathcal{F}, H', t \models_{\mathcal{I}} \phi$ for
 some history H' such that $t \in H'$
 $\mathcal{F}, H, t \models_{\mathcal{I}} \Box\phi$ if and only if $\mathcal{F}, H', t \models_{\mathcal{I}} \phi$ for
 every history H' such that $t \in H'$

In these rules, because t is in both H and H' , and branching is only into the future, these histories must coincide at least up to t . We can now interpret compound formulae such as

$\Diamond F\phi$ ϕ will be true in some possible future.
 $\Box F\phi$ ϕ will be true in every possible future.
 $P\Box\phi$ It was inevitable that ϕ would eventually be true.

Interval Semantics

In the standard semantics for Tense Logic, the elements of a temporal frame are instants, that is,

temporal elements lacking duration. A further possibility is to use temporal frames containing intervals as well as, or instead of, instants. This approach has been favored by those seeking to give a logical account of aspect in natural language, which requires us to handle events that take time (Dowty, 1979). The statement *John runs a mile* ($=\phi$) reports the occurrence of an event of several minutes' duration. It seems inappropriate to evaluate ϕ at instants, because there is no instant at which John runs a mile. Instead, interval semantics evaluates ϕ as true with respect to any interval i over which John runs a mile. Because running a mile is an accomplishment (Vendler, 1967), this means that ϕ will not be true on any proper subinterval of i —in contrast to the progressive form *John is running a mile*, which would be true.

Formally, we use frames of form $\mathcal{F} = (I, \prec, \sqsubset)$, where I is a set of intervals, \prec is the 'earlier than' relation on intervals, and \sqsubset is the proper subinterval relation. Accomplishments and activities are distinguished by their different entailments: an accomplishment sentence satisfies the rule

(ACC) If $\mathcal{F}, i \models_{\mathcal{I}} \phi$ and $i' \sqsubset i$ then $\mathcal{F}, i' \not\models_{\mathcal{I}} \phi$,

whereas an activity sentence satisfies

(ACT) If $\mathcal{F}, i \models_{\mathcal{I}} \phi$ and $i' \sqsubset i$ then $\mathcal{F}, i' \models_{\mathcal{I}} \phi$.

The progressive aspect might be regarded as an operator *Prog* for (among other things) converting accomplishments into activities. The following semantic rule for *Prog* is plausible:

$\mathcal{F}, i \models_{\mathcal{I}} \text{Prog}(\phi)$ if and only if $\mathcal{F}, i' \models_{\mathcal{I}} \phi$ for some i' such that $i \sqsubseteq i'$,

and this rule has the consequence that if ϕ is an accomplishment (as defined by ACC) then *Prog*(ϕ) is indeed an activity (as defined by ACT). Thus, basic Tense Logic can be enhanced to capture some of the complexities of temporality in natural language.

This analysis only covers cases where *Prog*(ϕ) is true by virtue of an eventually completed occurrence of the event expressed by ϕ . By the so-called 'imperfective paradox', the former can be true without the latter ever becoming true (e.g., if John stops running before the mile is completed). This property can be captured by combining interval semantics with branching time. A history is now a maximal linear set of intervals (where linearity of a set of intervals can be defined in terms of \prec and \sqsubset). A more accurate rendition of the progressive is now

$\mathcal{F}, H, i \models_{\mathcal{I}} \text{Prog}(\phi)$ if and only if $\mathcal{F}, H', i' \models_{\mathcal{I}} \phi$ for some H, i' such that $i \sqsubseteq i' \in H'$.

Because i is in both H and H' , and branching is only into the future, H and H' must coincide at least up to

the end of i ; thereafter they may – but need not – diverge, allowing for the possibility that the completion of ϕ occurs in a different history from the one in which *Prog*(ϕ) is being evaluated.

Other Forms of Temporal Logic

The systems discussed above may be thought of as variants of Prior's basic Tense Logic, expressing temporal properties and relations by means of operators. Alongside these, there has been a tradition of modeling temporal information in first-order logic, using individual terms as explicit bearers of temporal reference. Whereas classically one might write, say, *Asleep(john)* for *John is asleep*, the *Method of Temporal Arguments* writes *Asleep(john, t)* to say that John is asleep at time t . Tenses are paraphrased using constructions that mirror the semantic rules for tense operators, e.g., $\exists t' (t' \prec t \wedge \text{Asleep}(\text{john}, t'))$ says that 'John has been asleep' holds at t . For events that take time, we use interval arguments, e.g., *Run(john, mile, i)*, where s is the interval over which John runs a mile.

A well-known variant of this method, from Donald Davidson (1967), replaces temporal arguments by terms denoting individual events, e.g., $\exists e (\text{Run}(\text{john}, \text{mile}, e) \wedge t' \text{Occurs}(e, i))$, which says that some event e is the running of a mile by John and occurs on interval i . An advantage of this is that it allows us to analyze adverbial modifiers by means of additional event predicates linked by logical conjunction, e.g., $\exists e (\text{Run}(\text{john}, \text{mile}, e) \wedge \text{Occurs}(e, i) \wedge \text{In}(e, \text{field}))$.

Davidson's method is sometimes known as 'event-token reification', because the e terms refer to *event tokens*, i.e., individual occurrences. An alternative form of reification, particularly favored in Artificial Intelligence, involves state and event *types* (Allen, 1984). In this method, *run(john, mile)* is a *term* denoting the event type 'John runs a mile'; an occurrence of this type on interval i is expressed as *Occurs(run(john, mile), i)*. A separate predicate is used for state-types, e.g., *Holds(asleep(john), i)*, meaning that John is asleep *throughout* interval i . An advantage claimed for this style of analysis is that by allowing quantification over types it enables expression of general temporal knowledge concerning, for example, causality, e.g., $\forall e, s, e', d (\text{Cause}(e, s, e', d) \wedge \text{Holds}(s, t) \wedge \text{Occurs}(e, t) \rightarrow \text{Occurs}(e', t + d))$. The first conjunct states that events of type e occurring under condition s give rise to events of type e' after a delay of length d .

See also: Event-Based Semantics; Modal Logic.

Bibliography

- Allen J F (1984). 'Towards a general theory of action and time.' *Artificial Intelligence* 23, 123–154.
- Davidson D (1967). 'The logical form of action sentences.' In Rescher N (ed.) *The logic of decision and action*. Pittsburgh: University of Pittsburgh Press. 81–95. [Repr. in Davidson D (1980). *Essays on actions and events*. Oxford: Clarendon Press. 105–122].
- Dowty D R (1979). *Word meaning and Montague Grammar*. Dordrecht: D. Reidel.
- Øhrstrøm P & Hasle P F V (1995). *Temporal logic: from ancient ideas to artificial intelligence*. Dordrecht: Kluwer Academic Publishers.

- Prior A N (1957). *Time and modality*. Oxford: Clarendon Press.
- Prior A N (1967). *Past, present and future*. Oxford: Clarendon Press.
- Prior A N (2003). *Papers on time and tense*. Hasle P, Øhrstrøm P, Bräuner T & Copeland J (eds.). Oxford: Oxford University Press.
- Van Benthem J (1991). *The logic of time: a model-theoretic investigation into the varieties of temporal ontology and temporal discourse* (2nd edn.). Dordrecht: Kluwer Academic Publishers.
- Vendler Z (1967). *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

Tense and Time: Philosophical Aspects

L N Oaklander, University of Michigan–Flint, Flint, MI, USA

© 2006 Elsevier Ltd. All rights reserved.

As we ordinarily think and talk about time, it is a truism that time passes. Dates, like the events that occur at those dates, are once in the future, then become present, and then recede into the more and more distant past with the passage of time. To think of time as passing, and events as changing with respect to the characteristics of pastness, presentness, and futurity, is to conceive of the transient aspect of time or temporal becoming. The most fundamental debate in the philosophy of time concerns the status of temporal becoming. Do events **really** pass from the future to the present and into the past, as tensed- or A-theorists (derived from McTaggart's 1908, 1927 A-series of past, present, and future moments or events) such as C. D. Broad (1923, 1938), Richard Gale (1968), A. N. Prior (1967), George Schlesinger (1980), Quentin Smith (1993), Storrs McCall (1994), Michael Tooley (1997), William Lane Craig (2000a, 2000b), and others have maintained? Or is the passage of time a myth and an illusion, as B-theorists (derived from McTaggart's B-series of events ordered by the relations earlier, later, and simultaneous) such as Bertrand Russell (1915), J. J. C. Smart (1980), Robin Le Poidevin (1991), D. H. Mellor (1998), L. Nathan Oaklander (2004b) and others have maintained? That is one issue. Another closely connected issue concerns the proper analysis of tense in ordinary language and thought. We express the passage of time (or the myth of passage) by means of tensed discourse and tensed beliefs. For example, we ordinarily say, at different times, that an event will occur, is occurring,

and did occur, and it is commonplace to believe that, for example, today is Monday, tomorrow will be Tuesday, and yesterday was Sunday. Two questions of analysis concerning these ordinary tensed sentences and beliefs immediately arise: (1) What is the meaning of tensed discourse? and (2) what are the truth conditions or truthmakers of tensed sentences? A third issue, intimately related to the other two, concerns the reference of temporal indexicals (such as now, 'yesterday,' and 'tomorrow'). Do temporal indexicals refer directly to some items (such as times or sentence tokens); or do they refer indirectly to items via a mediating sense (such as the property of **presentness**); or do they, perhaps, perform both functions, or neither?

Although these issues are fundamentally metaphysical in nature, the dispute between the A- and B-theorists has, until the 1980s, centered on temporal language. Defenders of the B-view have often argued that since tensed discourse could be eliminated or translated without loss of meaning into tenseless discourse, an adequate account of the nature of time need not countenance any special kind of tensed fact or tensed properties. In other words, the old B-theory of time assumed that a logical analysis of ordinary language that eliminates tensed discourse supported an ontological analysis of time that rejects transient temporal properties of **pastness**, **presentness**, and **future**, or any other form of passage. The A-theorist shared that assumption, but argued that since no tenseless translations were successful, temporal becoming in some form or another is necessary in any adequate account of time. A-theorists claim, in other words, that because tensed discourse is ineliminable, the B-theorist is mistaken and tensed properties and facts must exist.

For a variety of reasons, some having to do with arguments in the philosophy of time and some having to do with arguments in the philosophy of language, recent defenders of the B-theory have come to embrace the thesis that tensed sentences cannot be translated by tenseless ones without loss of meaning. The reason for this is clear. It is part of the meaning of sentences reflecting temporal becoming that they change their truth value with the passage of time. For example, a token (or instance) of the sentence type 'I will be celebrating my 60th birthday in the future' is true today, on January 7, 2005, but in six months another token of that sentence will be false. On the other hand, it is part of the meaning of tenseless sentences expressing temporal relations between events that different tokens of the same tenseless sentence have the same truth value whenever they are expressed. Thus, for example, the linguistic meaning of the sentence (S) 'My 60th birthday is future' cannot be captured by (V) 'My 60th birthday is later than January 7, 2005' even if (S) is uttered on January 7, 2005, since (V) is always true, whereas on July 7, 2005 (S) is not.

Nevertheless, recent B-theorists have denied that the ineliminability of tensed language and thought entails the reality of temporal properties or temporal becoming. According to the new B-theory of time, our need to think and talk in tensed terms is perfectly consistent with its being the case that time itself is tenseless. Tensed discourse is indeed necessary for timely action, but tensed facts are not, since the truth conditions of tensed sentences can be expressed in a tenseless metalanguage that describes unchanging temporal relations between and among events.

There are two popular versions of the new B-theory of time, the token-reflexive account (Mellor, 1981) and the date-analysis account (Smart, 1980). On the token-reflexive account the temporal relation between the date at which one says, thinks, or writes down a tensed sentence and the event or thing that it is about provides an objective basis for the truth-value of any tensed sentence. A present-tense sentence token is true if, and only if, it occurs (exists tenselessly) at (roughly) the same time as the event it is about; a past-tense token is true if, and only if, it occurs at a time later than the event it refers to, and so on. Thus, on the token-reflexive account the truth conditions of tensed sentence and judgment tokens are tenseless facts – that is, facts that involve only temporal relations between and among nontensed events.

The date-analysis version of the new B-theory also denies the thesis of linguistic reducibility and claims instead that corresponding to every tensed sentence-token is a tenseless sentence that gives its truth conditions. For example, J. J. C. Smart (1980) claims that

the notion of becoming present seems a pretty empty notion, and this is even more obvious when we recognize the indexical nature of words like 'present', 'past' and 'future'. When a person P utters at a time *t* the sentence 'Event E is present' his assertion is true if, and only if, E is at *t*. More trivially, when P says at *t* 'time *t* is now' his assertion is true if, and only if, *t* is at *t* so that if P says at *t* 't is now' his assertion is thereby true" (1980: 11).

The heart of the date-analysis thesis is that temporal indexicals like 'now,' 'this time,' and 'the present,' as used on a given occasion, are referring terms that denote a time. Thus, if a temporal indexical sentence such as 'Event E is now occurring' is uttered at *t*₁, then it reports an event that is identical with the event reported at any time by the use of the nonindexical sentence 'Event E is occurring at *t*₁.' On this view, indexicals and proper names such as dates are rigid designators (Kripke, 1980). Thus, a tensed sentence like 'It is now 1980,' uttered in 1980, reports the same fact as the necessary truth reported by 'It is 1980 in 1980' or '1980 is at 1980.' It does not follow, nor is it part of the date-analysis thesis to maintain, that 'It is now 1980' and '1980 is at 1980' express the same proposition or have the same meaning (Beer, 1988). Nevertheless, to know that the truth condition of a tensed sentence token is a function of the date of its occurrence, and what it is about, is to know the meaning of a tensed sentence (Mellor, 1998).

Both versions of the new B-theory have been the subject of considerable debate. Proponents of the A-theory have argued that the B-theory cannot give an adequate account of the truth conditions of A-sentences, whereas B-theorists have denied this (see Oaklander and Smith, 1994, Part I; Le Poidevin, 1998; Smith, 1999; Craig, 2000b; Mozerky, 2000; and Dyke, 2002, 2003a). More recently, Oaklander (2004a) has argued that tensed sentence-tokens and the mind-independent contents they express, while pragmatically useful and perhaps indispensable, are strictly and metaphysically false and couldn't possibly be true since McTaggart's paradox (to be considered shortly) has demonstrated that the A-theory of time is contradictory.

In summary, the issue that now rages between the various camps in the A/B-theory debate concerns the truth conditions of statements that reflect temporal becoming and temporal relations. In this context, truth conditions are truthmakers: the basis in reality for those true sentences that record facts about the transitory and temporal relational aspects of time. Thus, the A/B debate centers on the questions: What, if anything, do the tenses and our use of temporal concepts reflect about the metaphysical nature of time? And, What is the ultimate metaphysical foundation of our experience of succession and

temporal relations? To deal with these questions it is necessary to consider McTaggart's paradox.

McTaggart maintained that becoming is an essential feature of temporal reality, since without real passage there is no change, and without change there is no (B-) time. If events simply stand in the unchanging relations of earlier/later than, nothing would first have a property and then lose it. Thus, in order for there to be change, it must be the case that events have the properties of pastness, presentness, and futurity and change with respect to those determinations. McTaggart argued, however, that such change was impossible. For if every event is past, present, and future, then since A-characteristics are incompatible, we arrive at a contradiction. Of course, the seemingly obvious way out of the 'paradox' is to appeal to tense: it is not the case that every event is (simultaneously or tenselessly) past, present, and future, but rather every event either is past, was future and present; or is present, will be past, and was future; or is future and will be past and present. This way of avoiding the paradox is well taken, but it certainly is not the last word, since we need to understand how the tenses 'is now,' 'was,' and 'will be' are to be interpreted if we are to understand how such an account is to render temporal becoming possible.

A currently fashionable A-view of the tenses is called 'presentism.' According to this view, only the present exists; the past and future are species of unreality, that is, the past and future tenses are operators (with no ontological significance) on present tense sentences (Prior, 1967; Hinchliff, 1996). Given that only the present exists, there is no longer any problem of how one and the same event (or moment) can have incompatible A-determinations, since there is only one A-characteristic, namely presentness.

Nevertheless, presentism is not without its difficulties. To see why, suppose we start with

FPe & Ne & PFe

This is read as 'e will be past' and 'e is present' and 'e was future.' Perhaps there is no contradiction in such a representation of becoming, but we are still left with the question: What are the truthmakers for the first and last conjuncts? More specifically, what is the ontological difference between *FPe* and *PFe*, given that neither 'F' nor 'P' is a predicate that ascribes properties to *e*? Without such an account, the appeal to grammatically consistent tensed statements is a vacuous response to McTaggart's paradox.

The B-theory, in rejecting temporal passage as an objective feature of reality over and above the tenseless existence of events spread along the four-dimensional manifold, avoids McTaggart's conundrum, but it has questions of its own to answer.

If events don't really pass, then why do we talk as if they do? And if nothing is really present, then what account can be given on the presence of experience, and our experience of the present? (For a discussion of proposed solutions to these questions see Oaklander and Smith, 1994, Part III; Dyke and Maclaurin, 2003; and Oaklander, 2004b, Part III).

See also: Anaphora: Philosophical Aspects; Essential Indexical; Indexicality: Philosophical Aspects; Type versus Token.

Bibliography

- Beer M (1988). 'Temporal indexicals and the passage of time.' *The Philosophical Quarterly* 38, 158–164. [Reprinted in Oaklander & Smith (eds.), 87–93.]
- Broad C D (1923). *Scientific thought*. London: Routledge & Kegan Paul. [Reprinted 1969, New York: Humanities Press.]
- Broad C D (1938). *Examination of McTaggart's philosophy* 2. Cambridge: Cambridge University Press. [Reprinted 1976, New York: Octagon Books.]
- Craig W L (2000a). *Tensed theory of time: a critical examination*. Dordrecht: Kluwer Academic Publishers.
- Craig W L (2000b). *The tenseless theory of time: a critical examination*. Dordrecht: Kluwer Academic Publishers.
- Dyke H (2002). 'Tokens, dates and tenseless truth conditions.' *Synthese* 131, 329–351.
- Dyke H (2003). 'Tensed meaning: a tenseless account.' *Journal of Philosophical Research* 27, 67–83.
- Dyke H & Maclaurin J (2003). "'Thank goodness that's over": the evolutionary story.' *Ratio* XV, 276–292.
- Gale R (1968). *The language of time*. New York: Routledge & Kegan Paul.
- Hinchliff M (1996). 'The puzzle of change.' In Tomberlin J (ed.) *Philosophical Perspectives 10: Metaphysics*. Cambridge, MA: Blackwell. 119–136.
- Jokic A & Smith Q (eds.) (2003). *Time, tense and reference*. Cambridge: MIT Press.
- Kripke S (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- LePoidevin R (1991). *Change, cause and contradiction: a defense of the tenseless theory of time*. New York: St. Martin's Press.
- LePoidevin R (1998). 'The past, present, and future of the debate about tense.' In Le Poidevin R (ed.) *Questions of time and tense*. Oxford: Clarendon Press. 15–42.
- McCall S (1994). *A model of the universe*. Oxford: Clarendon Press.
- McTaggart J E M (1927). 'Time.' In Broad C D (ed.) *The nature of existence* 2. Cambridge: Cambridge University Press. [Reprinted 1968, Grosse Pointe, MI: Scholarly Press, 9–31.]
- McTaggart J E M (1908, 1934). 'The unreality of time.' *Mind* 18, 457–474. In Keeling S J (ed.), *Philosophical studies*. London: Edward & Arnold & Co. 110–134.

- Mellor D H (1981). *Real time*. Cambridge: Cambridge University Press.
- Mellor D H (1998). *Real time II*. London: Routledge.
- Mozersky J (2000). 'Smith on times and tokens.' *Synthese* 129, 405–411.
- Oaklander L N (2004a). 'Two versions of the new B-theory of language.' In Oaklander L N (eds.). (2004b), 265–292.
- Oaklander L N (2004b). *The ontology of time. Studies in analytic philosophy*. Series editor, Quentin Smith Amherst, NY: Prometheus Books.
- Oaklander L N & Smith Q (eds.) (1994). *The new theory of time*. New Haven, CT: Yale University Press.
- Prior A N (1967). *Papers on time and tense*. Oxford: Oxford University Press.
- Russell B (1915). 'On the experience of time.' *Monist* 25, 212–233.
- Schlesinger G (1980). *Aspects of time*. Indianapolis, IN: Hackett Publishing Co.
- Smart J J C (1980). 'Time and becoming.' In van Inwagen P (ed.) *Time and cause: essays in honor of Richard Taylor*. Boston: Reidel. 3–15.
- Smith Q (1993). *Language and time*. New York: Oxford University Press.
- Smith Q (1999). 'The "sentence-type version" of the tenseless theory of time.' *Synthese* 119, 233–251.
- Smith Q (2002). 'Time and degrees of existence.' In Callender C (ed.) *Time, reality, and experience*. Cambridge: Cambridge University Press. 119–136.
- Tooley M (1997). *Time, tense, and causation*. Oxford: Clarendon Press.

Testimony

D Matheson, Carleton University, Ottawa, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

Testimony as a Belief Source

A belief source is a cognitive process that takes certain characteristic mental states as input and yields the representational state of belief as output. In the broadest sense of interest to epistemologists, testimony may be understood as a belief source whose input and output consist, respectively, of (a) reception of another's report about the world and (b) belief in what is reported. As such, testimony has a wide variety of manifestations, ranging from the highly formal to the exceedingly mundane – from taking the word of a sworn witness in a legal trial to accepting the nutritional information on a box of cereal.

Reflection on the ubiquity of testimony in this broad sense highlights the importance of language for our representation of the world. Thus, compare testimony with such belief sources as sensory perception, inference, and retentive (as opposed to recollective) memory, which may be characterized roughly as in the first three rows of **Table 1**.

In the case of the nontestimonial belief sources, description of the salient input need not explicitly mention language. Regarding testimony as a belief source, however, the appeal to language is unavoidable. Reception of another's report about the world quite obviously involves some kind of grasp of another's linguistic act – another's saying that things are thus-and-so, for example – or the linguistic effects

thereof, such as the written record of another's say-so. Simply put, much of what we believe about the world would be lost without testimony, and without language, testimony would be lost.

Testimony as a Knowledge Source

Reflection on the ubiquity of testimony also highlights the potential importance of language for our felicitous representation of the world. To many epistemologists, testimony plays a central role in affording us not only mere beliefs about but also knowledge of the way things are, where knowledge is distinguished from mere belief at least partly by virtue of requiring justification or derivation from a proper source. (No matter how firmly one believes that supernatural phenomena exist, for example, one's belief that they do fails to amount to knowledge if it is unjustified or derived from an unreliable belief source such as wishful thinking.) As social creatures, Reid (1764/1997: 193) claimed, "We receive the greatest and most important part of our knowledge by the information of others." Echoing the point, Chakrabarti (1994b: 965) notes that "[w]itnesses in law courts, trusted travelers, historians, scientists, dictionaries, radio news-readers, parents, and computers have told us that something is the case and consequently we have come to *know* that it is so."

Skepticism

This common view is by no means universally held. Locke (1689/1989: 58) considered that "[t]he floating of other men's opinions in our brains makes us not

one jot the more knowing, though they happen to be true. What in them was science is in us opiniatrety.” Similarly, Barnes (1980: 200), who acknowledges that “we all do pick up beliefs in [a] second hand fashion,” expresses the “fear that we often suppose that such scavenging amounts to knowledge.” That, he claims, “is only a sign of our colossal credulity: [testimony] is a rotten way of acquiring beliefs, and it is no way at all of acquiring knowledge.”

Skeptics such as Locke and Barnes insist that the most we can get when it comes to the connection between testimony and our representation of the world is the last row of Table 1: Testimony, in their view, can at best generate mere beliefs about the world – that is, beliefs that fail the justification or derivation-from-a-proper-source requirement on knowledge. Others, such as Reid and Chakrabarti, insist that the last row frequently describes a knowledge source: for them, what falls in the output box is often new knowledge of the world.

Even if we grant the importance of testimony for a felicitous representation of the world, we may well wonder how it works. Here, there are three questions to be distinguished. First, how more fully are we to understand the salient input to testimony if it is indeed to be understood as a knowledge source? Second, what, if anything, must be added to the salient input to get the total input to testimony as a knowledge source? Third, how are we to understand the cognitive process that leads from the total input to new knowledge of the world?

Perception and Understanding

Regarding the first question, there is little doubt that sensory perception is involved: In receiving another’s report about the way things are, one comes to know that another has reported that things are thus-and-so on the basis of visual, auditory, etc. experience of the physical items that encode the report. However, bare sensory perception does not suffice. Both the monolingual Japanese hearer and the Anglophone hearer might equally perceive through their senses a speaker’s auditory deliverance ‘I’m

tired today,’ even though only the latter may count as receiving the speaker’s report that he or she is tired today, because only the latter may understand the meaning of the perceived deliverance. Generally, then, the salient input to testimony as a knowledge source amounts to knowledge, acquired through sensory perception of an understood linguistic act (or its effects), that another has reported that things are thus-and-so.

Reductionism and Anti-reductionism

The second and third questions bring us to the heart of the most prominent issue in epistemologists’ attempt to understand the nature of testimony as a knowledge source, viz. the reductionist/anti-reductionist debate. Whereas reductionists (Fricker, 1987; Hume, 1777/1995) maintain that testimony as a knowledge source reduces across the board to a species of inference, anti-reductionists (Burge, 1993; Chakrabarti, 1994a; Coady, 1992; McDowell, 1994; Reid, 1764/1997; Strawson, 1994) demur.

The kind of inference to which testimony as a knowledge source reduces, according to reductionists, is inference from the salient input together with knowledge of the trustworthiness of the reporter with respect to what he or she has reported. Hence, reductionists’ answer to the second question is that knowledge of the trustworthiness of the reporter must be added to the salient input to get the total input to testimony as a knowledge source. Their answer to the third question is that we are to understand the cognitive process leading from testimony’s total input to its output as one of inference. These answers jointly present a model of testimony as a knowledge source that is roughly captured in Table 2.

Anti-reductionists’ answer the second question, in contrast, holds that in many cases nothing must be added to the salient input to get the total input to testimony as a knowledge source. Their response to the third question is that the cognitive process leading from testimony’s total input to its output is better thought of simply along the lines of causation in general, since inference in particular is not required

Table 1 Belief sources

<i>Salient input</i>	<i>Cognitive process</i>	<i>Output</i>
<i>Sensory experience of the world</i> , e.g., having the visual impression of a chair over yonder	<i>Sensory perception as a belief source</i>	<i>New belief about the world</i> , e.g., believing that there is a chair over yonder
<i>Old belief about the world</i> , e.g., believing that Socrates is a man and that all men are mortal	<i>Inference as a belief source</i>	<i>New belief about the world</i> , e.g., believing that Socrates is mortal
<i>Old belief about the world</i> , e.g., believing that it was cold on Tuesday	<i>Retentive memory as a belief source</i>	<i>Old belief about the world</i> , e.g., believing that it was cold on Tuesday
<i>Reception of another’s report about the world</i> , e.g., hearing another say that it’s raining	<i>Testimony as a belief source</i>	<i>New belief about the world</i> , e.g., believing that it’s raining

Table 2 Reductionist model of testimony as a knowledge source

Input	Cognitive process	Output
(a) Knowledge, acquired through sensory perception of understood linguistic acts (or their effects), that another has reported something about the world, and (b) knowledge that the reporter is trustworthy with respect to his or her report about the world	Inference from input to output	New knowledge of the world

Table 3 Anti-reductionist model of testimony as a knowledge source

Input	Cognitive process	Output
Knowledge, acquired through sensory perception of understood linguistic acts (or their effects), that another has reported something about the world	Causation of output by input	New knowledge of the world

(although it may in some cases be present). Thus, the anti-reductionist opts for something like the model given in Table 3.

(Stepping outside of the recipient's cognitive processes, it should be noted, the anti-reductionist requires certain external conditions to be satisfied for the output of testimony as a knowledge source to amount to knowledge. Like the reductionist, he or she requires that what is reported be true; he or she further requires that the reporter in fact be trustworthy with respect to his or her report, even though the recipient need not know that he or she is trustworthy.)

Reductionists claim that the anti-reductionist model lends itself all too easily to something like epistemic gullibility and hence fails to serve as a plausible model for testimony as a legitimate knowledge source. Anti-reductionists point to cases in which it is supposedly clear that the reductionist model does not apply, most notably those involving young children's acquisition of linguistic knowledge. (Such acquisition seems to be in large part testimonial, but, say anti-reductionists, it cannot with any plausibility be regarded as typically involving inference from knowledge of the trustworthiness of reporters.)

Considerations of this sort have tended to shift the focus of the reductionist/anti-reductionist debate to the question of how to understand testimony as a knowledge source for mature cognitive agents, already possessed of the sort of linguistic knowledge just mentioned (Fricker, 1994, 1995).

Summary

Testimony is undeniably a major source of our beliefs about the world, a fact that underscores the centrality of language for our representational capacities. Many maintain that testimony is also a major source of our knowledge of the world, a contention that implies the centrality of language for our felicitous representational capacities. Even granting this contention, however, there is considerable debate about the nature of testimony. Clearly, it involves knowledge, through sensory perception of understood linguistic acts (or their effects), that another has reported that things are thus-and-so. In the reductionist model, it reduces across the board to inference from such knowledge and from knowledge of the reporter's trustworthiness with respect to his or her report. In the anti-reductionist model, no such reduction obtains.

See also: Cognitive Science and Philosophy of Language; Communication, Understanding, and Interpretation: Philosophical Aspects; Epistemology and Language; Representation in Language and Mind; Thought and Language: Philosophical Aspects.

Bibliography

- Barnes J (1980). 'Socrates: the jury, Part II.' *Proceedings of the Aristotelian Society, Supplementary Volume* 54, 193–206.
- Burge T (1993). 'Content preservation.' *Philosophical Review* 107, 457–488.
- Chakrabarti A (1994a). 'Telling as letting know.' In Chakrabarti & Matilal (eds.), 99–124.
- Chakrabarti A (1994b). 'Testimony: a philosophical study [Review essays].' *Philosophy and Phenomenological Research* 54, 965–972.
- Chakrabarti A & Matilal B K (eds.) (1994). *Knowing from words: Western and Indian philosophical analysis of understanding and testimony*. Dordrecht, The Netherlands: Kluwer.
- Coady C A J (1992). *Testimony: a philosophical study*. Oxford: Clarendon.
- Fricker E (1987). 'The epistemology of testimony.' *Proceedings of the Aristotelian Society, Supplementary Volume* 61, 57–83.
- Fricker E (1994). 'Against gullibility.' In Chakrabarti & Matilal (eds.), 125–162.

- Fricker E (1995). 'Telling and trusting: reductionism and anti-reductionism in the epistemology of testimony.' *Mind* 104, 393–411.
- Hume D (1995). 'An enquiry concerning human understanding.' [Section X.] In Nidditch P H (ed.) *Enquiries concerning human understanding and concerning the principles of morals*, 3rd edn. Oxford: Clarendon. (Original work published 1777.)
- Locke J (1989). *An essay concerning human understanding*. Nidditch P H (ed.). Oxford: Clarendon. (Original work published 1689.)
- McDowell J (1994). 'Knowledge by hearsay.' In Chakrabarti & Matilal (eds.). 195–224.
- Reid T (1997). *An enquiry into the human mind on the principles of common sense*. Brookes D (ed.). University Park: Pennsylvania State University Press. (Original work published 1764.)
- Strawson P F (1994). 'Knowing from words.' In Chakrabarti & Matilal (eds.). 23–27.

Thought and Language: Philosophical Aspects

E Eaker, University of Western Ontario, London, Ontario, Canada

© 2006 Elsevier Ltd. All rights reserved.

There is a veritable thicket of philosophical issues waiting to ensnare anyone who reflects on the relation between thought and language. The complexity of this thicket has been further enhanced by the fact that many philosophers have been preoccupied with trying to understand the nature and limits of human thought about the world we inhabit. In fact, philosophical interest in the relation of thought to the world has, to a large extent, motivated philosophical interest in language. For example, consider the proposal that our concepts and the basic structural features of our thoughts reflect the fundamental ontological categories 'out there' in the world. Some philosophers have proposed that by a careful study of language we can obtain a clearer grasp of the structure of thought and therefore a clearer understanding of what the world is like. Others have argued that the last place to look for metaphysical insight is language, thinking it to be a distorting mirror of the world. In addition to these broader metaphysical and epistemological questions – i.e., questions concerning what the world is like and what we can know about it – philosophical interest in language has also been motivated simply by an understandable fascination with the amazing human capacities for thought and language. It has been emphasized time and again that these capacities set human beings apart from the rest of the natural world.

Two of the main philosophical questions that have arisen regarding the relation of language and thought will be discussed here. First, philosophers have questioned the relative priority of thought and language: is thought prior to language (in some important sense

to be defined) or vice versa? How one conceives of the relative priority of thought and language can affect how one approaches the second question: language and thought are both characterized by intentionality or 'aboutness'. When we think and speak, we are connecting with things in the world, and thinking and talking about them. Philosophers have sought to understand this aboutness of thought and language.

The Relative Priority of Thought and Language

Is thought prior to language? That is, is language merely a vehicle by means of which we express our thoughts? Some philosophers have argued (or merely assumed) that the capacity for linguistic expression is a capacity that allows us to communicate our thoughts and that we have an antecedent grasp on those thoughts. If so, one wants to understand what makes the structure of the natural languages well suited to expressing our thoughts. Are thoughts structured in a way similar to the languages that we use to express them? Is language a window onto the mind, so to speak?

Some have answered this last question affirmatively, but think that the reason the structure of human languages reflects the structure of human thought is because the capacity for language is **prior** to thought in some important sense. An extreme version of this view has been held by philosophers who are skeptical about the very existence of mental phenomena, such as thought. They point out that our tendency to attribute thoughts to other individuals depends largely on their capacity to use language. They argue that perhaps this linguistic behavior is all there really is to thought, properly so-called. Some have found this position too extreme, but have nonetheless stressed that thought, properly

so-called, is not possible without the acquisition of a language and that the idea that language merely serves as the vehicle of thought is misguided. I describe a classic 'Cartesian' view of the primacy of thought over language and briefly sketch a thought-skeptical 'behaviorist' response to it, followed by a more nuanced view that recognizes the existence of thoughts, but argues that language is still prior to thought in an important sense. These issues are very much live ones in current research in linguistics, psychology, and philosophy departments, but to avoid some of the complexity and technical vocabulary of more current discussions, I present more classic loci of the above three positions.

The Cartesian View

The following passages from Descartes (1596–1650) represent commonly held views on thought and the relation between thought and language in the early modern and modern period. First, Descartes exhibits the perennial fascination with thought, and by extension language, as that which distinguishes man from beast. Discussing the possibility of automata constructed to look like and mimic the movements of animals and humans, Descartes writes:

[I]f any such machines had the organs and outward shape of a monkey or of some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals; whereas if any such machines bore a resemblance to our bodies and imitated our actions as closely as possible for all practical purposes, we should still have two very certain means of recognizing that they were not real men. The first is that they could never use words, or put together other signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words which correspond to bodily actions causing a change in its organs (e.g., if you touch it in one spot it asks you what you want of it, if you touch it in another it cries out that you are hurting it, and so on.) But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dumbest of men can do. Secondly, even though such machines might do some things as well as we do them, or perhaps even better, they would inevitably fail in others, which would reveal that they were acting not through understanding but only from the disposition of their organs. (Descartes, 1999)

We see several themes in this passage. It is argued that what separates humans from the other animals is the capacity for thought and reason. This capacity underwrites the linguistic capacity of humans, whose

linguistic behavior cannot be adequately understood as the predetermined, mechanically generated output response to various external stimuli. To describe the human linguistic capacity in such a mechanical way would be to miss the evident fact that humans, but not automata (or nonhuman animals), understand to a significant degree the external stimuli, are able to reason to an appropriate response, and, on making an utterance, mean what they say. Human language is thus fundamentally rooted in human reason, which is in turn an operation of the human mind. For Descartes, the mind is a distinct substance from the body and its workings cannot be explained by the mechanistic laws that govern bodies. Since nonhuman animals have no minds, they can be described in completely mechanistic terms. How can we know this to be the case? Descartes continues:

Now in just these two ways we can also know the difference between man and beast. For it is quite remarkable that there are no men so dull-witted or stupid – and this includes even madmen – that they are incapable of arranging various words together and forming an utterance from them in order to make their thoughts understood; whereas there is no other animal, however perfect and well-endowed it may be, that can do the like. This does not happen because they lack the necessary organs, for we see that magpies and parrots can utter words as we do, and yet they cannot speak as we do: *that is, they cannot show that they are thinking what they are saying*. On the other hand, men born deaf and dumb, and thus deprived of speech-organs as much as the beasts or even more so, normally invent their own signs to make themselves understood by those who, being regularly in their company, have the time to learn their language. (Descartes, 1999)

Again, Descartes emphasizes the priority of thought to language when he characterizes language as essentially a vehicle of thought. What makes a string of noises a use of language is that it expresses what the speaker is thinking. Any real use of language requires that there be understanding, not mere parroting of sounds. Humans, but not magpies and parrots, produce words, and string them together into sentences in a certain way in an act intended to convey their thoughts. Mere strings of sounds, even word-like sounds, and sentence-like strings of them are meaningless unless produced with the intention of conveying thoughts. Descartes seems to presuppose that as humans – creatures with both mind and body – we are essentially endowed with the capacity for thought. If we are unable, for whatever reason, to learn or to produce the spoken or written language of our community, we will nonetheless have thoughts that we might wish to convey and will find some means for doing so.

Behaviorism

Some philosophers in the early 20th century questioned the rich ontology of mental phenomena that was taken for granted by Descartes. A particularly influential school of thought was 'behaviorism'. Not all behaviorists questioned the existence of unseen mental phenomena. Some were simply trying to demarcate what they took to be the appropriate data on which to base a science of psychology. Such 'methodological' behaviorists thought that scientific description of human psychology should focus exclusively on intersubjectively observable phenomena – in particular on behavior, including linguistic behavior. Such behaviorists had a proscription for scientific methodology in the field of psychology, but stopped short of metaphysical pronouncements. An 'ontological' behaviorist, on the other hand, is someone who wishes to give an analysis of mental vocabulary in terms of outwardly observable behavior, without reference to any phenomena that have those qualities thought to be characteristic of the 'mental' – e.g., being immaterial, inner, available as a result of immediate awareness or through 'introspection' and 'private' (i.e., inaccessible to someone else). An ontological behaviorist has a metaphysical bias against admitting the very existence of phenomena thus characterized and seeks to explain what mistakes we are making when we apparently make reference to such phenomena. Behaviorism as a methodological thesis is not necessarily committed to such a strong metaphysical bias. Nonetheless, for all behaviorists there is a strong sense in which language is prior to thought: either mental vocabulary is analyzed in terms of behavior, in particular linguistic behavior, or, in the scientific description of human beings, it is abandoned in favor of description of behavior.

Sellars: Language as a Precondition for Thought

Wilfred Sellars provides a more nuanced view, avoiding a direct confrontation with our rather strong intuitions that there are mental phenomena, while at the same time explaining why language is indeed a necessary precondition for thought and not the other way around. In *Empiricism and the philosophy of mind*, he defends the following revised version of the classical analysis of our common sense conception of thoughts:

[T]o each of us belongs a stream of episodes, not themselves immediate experiences, to which we have privileged, but by no means infallible, access. These episodes can occur without being 'expressed' by overt

verbal behavior, though verbal behavior is – in an important sense – their natural fruition. Again, we can 'hear ourselves think', but the verbal imagery that enables us to do this is no more the thinking itself than is the overt verbal behavior by which it is expressed and communicated to others. ...Now, it is my purpose to defend such a revised classical analysis of our common-sense conception of thoughts, and in the course of doing so I shall develop distinctions which will later contribute to a resolution, in principle, of the puzzle of *immediate experience*. But before I continue, let me hasten to add that it will turn out that the view I am about to expound could, with equal appropriateness, be represented as a modified form of the view that thoughts are *linguistic episodes*. (Sellars, 1997)

At the same time that Sellars wishes to defend a revised version of the view that there are 'inner' thought episodes, he also wishes to defend a revised version of the view that thoughts are linguistic episodes – the crude form of which is behaviorism. To explain what he means, Sellars introduces a piece of 'anthropological science fiction'. He imagines a stage of 'prehistory' in which humans are limited to using a 'Rylean language' – a language whose fundamental vocabulary is composed of terms for public properties and public objects located in space and enduring through time. It has resources for conjunction, disjunction, negation, and quantification; employs the subjunctive conditional; and allows for other 'looser' logical relations that are typical of ordinary discourse. Sellars asks: Would anything have to be added to the Rylean language to allow speakers of it to recognize and talk about themselves as beings that think (as well as having feelings, sensations, and perceptions)?

First, says Sellars, assume that they also have a way of making 'semantical' statements – i.e., statements about the meanings of the terms in their vocabulary, statements about what has been said by the assertion of a sentence, and a general ability to say that the assertion is true or false. If they have these resources, then they are arguably equipped to talk about thoughts – since thoughts share the same characteristics as bits of language. Just as we can say that when we use the expression 'tiger' we are talking about tigers, we can likewise claim to be thinking about tigers. Sellars proposes that we reverse the classical order of explanation and suppose "that the intentionality of *thoughts* can be traced to the application of semantical categories to overt verbal performances" (Sellars, 1997). Clearly, in order to support this proposal, Sellars must provide an account of the aboutness of linguistic expressions – what connects the term 'Tamela' with the individual, Tamela, whose name it is; what connects the term tiger with the

species *Felis tigris*, and by extension with individual tigers; what connects the term ‘hard’ with that property that nails, concrete, and marble have, but which cotton and rabbit fur lack. Also, Sellars must provide this account without making reference either to a speakers’ independent grasp in thought of some thing, kind, or property or to a speaker’s intentions to refer to such a thing so grasped. Note that from a discussion of the relative priority of thought and language, we have moved into a discussion of the intentionality of thought and language. Sellars’s view that language is prior to thought clearly constrains his explanation of the aboutness of thought and language.

Returning to Sellars’s explanation of how speakers might come to talk about inner episodes such as thoughts, recall that Sellars’s mythical Rylean language-speakers’ capabilities include a description of public objects, properties, and relations and the ability to discuss the semantic properties of words and sentences. At some point, they will notice that their fellows exhibit intelligent behavior even though they are not making any assertions. This will lead them, says Sellars, to develop a theory according to which there are inner episodes that are the beginning of a process that usually, though not always, culminates in overt verbal behavior. The model for these episodes will be overt speech and thus what this rudimentary psychological theory will posit will be ‘inner speech’. As it is inner speech, the semantical properties of outward speech will be applicable to it, and so the Ryleans can talk of the inner episodes as having meaning and as being about various public objects. The term ‘thoughts’ could be introduced to name these episodes of inner speech. Insofar as the Ryleans had good reason to posit these thoughts (on the basis of observing intelligent behavior that is performed in silence), it can be assumed that they really do exist, with a nature to be further determined as the ‘science’ develops.

Two questions immediately come to mind about Sellars’s account: One, if the overt verbal discourse is supposed on this theory to be the culmination of a process that begins with inner speech, i.e., thought, does this not take us back to the Cartesian view that overt linguistic behavior serves as a vehicle for the expression of thought? Two, if we are explaining the occurrence of overt linguistic behavior as the result of thoughts and we allow ourselves to talk about the semantic properties – the meaning and aboutness – of thoughts, are we not thereby analyzing the semantical properties of overt linguistic behavior in terms of the semantic properties of thought? In response to the first question, Sellars replies that, on the contrary, this theory “is perfectly compatible with

the idea that the **ability to have thoughts** is acquired in the process of acquiring overt speech and that only after overt speech is well established, can ‘inner speech’ occur without its overt culmination” (Sellars, 1997; boldface added). In response to the second question, he writes, “It must not be forgotten that the semantical characterization of overt verbal episodes is the primary use of semantical terms, and that overt linguistic events as semantically characterized are the model for the inner episodes introduced by the theory” (Sellars, 1997).

Sellars continues with a description of how we come to self-ascribe thoughts. Sellars proposes that a speaker of the Rylean language, S, can be ‘trained’ on the basis of observing his own behavior, to say of himself that he is thinking ‘p’, just as a third party might say on the basis of S’s behavior that S is thinking ‘p’. According to Sellars, speaker S can eventually “be trained to give reasonably reliable self-descriptions, using the language of the theory, without having to observe his [own] overt behavior.” With the spread of such reliable self-ascription of thoughts, the Ryleans will begin to talk of the ‘privileged access’ that each individual has to his or her own thoughts. Now, *prima facie* Sellars’s story might seem to be exactly backward: after all, do we not think of thoughts as the sorts of things that we simply reflect on, things we introspect, and things we know the contents of? Is it plausible that speakers learn a language that has evolved with the expressive power to allow third-person ascriptions of thought, and that speakers then, by extrapolation, extend those concepts to themselves, self-ascribing thoughts and eventually becoming authoritative reporters of their own thoughts?

The problem with this type of objection to Sellars’s myth is that it misses another key aspect of language, according to some philosophers, its essentially public and social nature. The public and social nature of language has been much emphasized by, e.g., Ludwig Wittgenstein, Paul Grice, and Saul Kripke. Sellars defends his picture by pointing out that it stresses the intersubjectivity of thought and the concepts that are used to describe such inner episodes, while not denying the fact that individuals do have privileged access to their thoughts and do make authoritative self-reports using these concepts. Of our use of concepts such as ‘think’, ‘believe’, and ‘fear’, in self-reporting, Sellars writes:

[T]he reporting role of these concepts – the fact that each of us has a privileged access to his thoughts – constitutes a dimension of the use of these concepts which is *built on* and *presupposes* this intersubjective status. My myth has shown that the fact that language is essentially an intersubjective achievement, and is learned in intersubjective

contexts – a fact rightly stressed in modern psychologies of language, thus by B. F. Skinner and by certain philosophers, e.g., Carnap, Wittgenstein – is compatible with the ‘privacy’ of inner episodes. It also makes clear that this privacy is not an ‘absolute privacy.’ For if it recognizes that these concepts have a reporting use in which one is not drawing inferences from behavioral evidence, it nevertheless insists that the fact that overt behavior *is* evidence for these episodes *is built into the very logic of these concepts*. . . (Sellars, 1997)

A Closer Look at the Relation between Thought and Language

We have seen only a sample of the various complex issues that arise when one considers the relative priority of language and thought. A few more questions might help to puncture the level of generality of the discussion thus far. Suppose with Sellars that mental concepts have “built into their very logic” that they are evidenced by overt behavior. What would this show about the primacy of language over thought? Sellars has told a tale in which thought emerges in individuals in a way that depends on their acquisition of a language. If he is right, then it would seem that he has given us a clear sense in which language is prior to thought. However, this is far from clear due to the sketchiness of Sellars’s tale. Work on this topic by Vygotsky and, more recently, Andy Clark has seen the more detailed development of the view that although we do not think in any particular language, the acquisition of a language provides the necessary scaffolding for thought. Some quite recent work by psychologists such as Elizabeth Spelke and Dedre Gentner also address these and related issues. (See, for example, Vygotsky, 1998, Clark, 2003 and Gentner and Goldin-Meadow, 2003.) Returning to Sellars, many questions remain untouched and suggest a much more complicated picture about the relation between thought and language, including the question of intentionality.

First, we might return to that part of Sellars’s myth in which language begins to emerge in a community. Plausibly, the initial steps will involve the introduction of names for particular things, categories of things, properties of things, and relations that they bear to one another. How can this practice emerge if there is not an antecedent awareness on the part of the protolanguage users of individuals as distinct from one another, an antecedent awareness of ways in which things resemble or fail to resemble one another, and so on? That an expression can be introduced for some individual, collection of individuals, kind of substance, or property of things would seem to require an antecedent focus on it. This would suggest

that the aboutness of linguistic items is not fundamental, but is instead derivative from the aboutness of such a prelinguistic focus of the mind.

Second, setting aside speculation into the origins of language, we can simply focus on individuals who are learning a language. In Sellars’s favor, it does seem obvious that learning a language greatly enriches an individual’s thought capacities. If I have no language, I am unlikely to have the concepts that would allow me to think: *The recent convictions of crooked CEOs who hold themselves above the law is symptomatic of the public’s dissatisfaction with the current economic climate*. Nonetheless, consider the true language novice: Perhaps it is necessary for learning terms such as ‘ball’ and ‘mommy’, that a child can already discern in a rudimentary way an object from its surroundings, can distinguish one object from another, can distinguish himself from his mother, and can tell his own feet from toys in his crib. Perhaps some rudimentary conceptual abilities are preconditions for the acquisition of language. These questions are the focus of extensive empirical study and there remains a considerable amount of controversy over the role of language in concept formation and development.

Third, leaving aside the language learner, we can focus on the linguistically competent adult who asserts: *The man in the corner wearing the toupee is Janice’s employee*. What makes it the case that this assertion is **about** some particular individual? Is it simply the fact that some man is standing in the corner indicated by the speaker and is the only such man wearing a toupee? In other words, does the definite description used, and the rules of language that govern how definite descriptions are to be used to refer to individuals by means of their properties, determine who is being referred to? Or are the intentions and mental focus of the speaker also important? Suppose there is a man in the corner but that he is not wearing a toupee. He is, nonetheless, the person to whom the speaker intends to refer. Can we make sense of the speaker having such an intention, independently of her use of a certain description? It would seem unintuitive to deny this. If we agree that it makes sense to say that the speaker **intends** to refer to a particular man, regardless of the description she uses, would we say that the speaker nonetheless **fails** to refer to that man, simply because that man does not fit the definite description *the man in the corner wearing the toupee*? In general, do linguistic expressions and the rules that govern them determine their own path to objects in the world or is the referential role of linguistic expressions parasitic on the more primitive aboutness of our thoughts?

All three of the above paragraphs concern the relative priority of thought and language when it comes to explaining the aboutness of our thoughts and of the 'referring expressions' in our language. But all three focus on a different stage in the development, acquisition, and use of language. Thus, not only the questions, but also the considerations that should be brought to bear in answering them will be different. Though this essay provides but a sampling of the variety and complexity of the issues concerning the relation between thought and language, it should help to explain why philosophers, linguists, and psychologists must spend as much time clearly articulating precisely what questions they are concerned with, as they do in looking for answers to them.

See also: Analytic Philosophy; Behaviorism: Varieties; Causal Theories of Reference and Meaning; Communication, Understanding, and Interpretation: Philosophical Aspects; Epistemology and Language; Externalism about Content; Intention and Semantics; Representation in Language and Mind.

Bibliography

- Burge, T (1979). 'Individualism and the mental.' *Midwest Studies in Philosophy* 4, 73–121.
- Chomsky N (1985). *Knowledge of language: Its nature, origin and use*. Westport, CT: Praeger.
- Clark A (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. New York: Oxford University Press.
- Descartes R (1999). 'Discourse on the method.' In *The philosophical writings of Descartes I*. Cottingham J, Stoothoff R & Murdoch D (trans.). Cambridge: Cambridge University Press. 139–141.
- Donnellan K (1966). 'Reference and definite descriptions.' *Philosophical Review* 75, 281–304.
- Fodor J (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Fodor J (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.
- Frege G (1956). 'The thought: A logical inquiry.' A M & Quinton M (trans.). *Mind* 65, 289–311.
- Frege G (1980). 'Ueber Sinn und Bedeutung.' In *Translations from the philosophical writings of Gottlob Frege*. Black M (trans.) and Geach P & Black M (eds.). Oxford: Basil Blackwell. 56–78.
- Grice H P (1975). 'Logic and conversation.' In Cole P & Morgan J L (eds.) *Syntax and semantics* 3. New York: Academic Press. 41–58.
- Kripke S (1977). 'Speaker's reference and semantic reference.' In French P A, Uehling T E Jr & Wettstein H K (eds.) *Studies in the philosophy of language*. Midwest Studies in Philosophy, No. 2. Minneapolis: University of Minnesota Press. 255–276.
- Kripke S (1980). *Naming and necessity*. Cambridge, UK: Harvard University Press.
- May R (1985). *Logical form: Its structure and derivation*. Cambridge, MA: MIT Press.
- Putnam H (1975). 'The meaning of "meaning".' In *Language, Mind and Knowledge*. Minnesota Studies on the Philosophy of Science, No. 7. Minneapolis: University of Minnesota Press.
- Russell B (1919). 'Descriptions.' In *Introduction to mathematical philosophy*. London: George Allen and Unwin. 167–180.
- Ryle, G. *The concept of mind*. Chicago: The University of Chicago Press.
- Sellars W (1997). *Empiricism and the philosophy of mind*. Cambridge, MA: Harvard University Press.
- Strawson P F (1950). 'On referring.' *Mind* 59, 320–344.
- Vygotsky L S (1998). *The collected works of L. S. Vygotsky*. New York: Plenum Press.
- Wittgenstein L (1958). *Philosophical investigations*, Anscombe G E M (trans.). Oxford: Basil Blackwell.
- Wittgenstein L (1929). 'Some remarks on logical form.' *Aristotelian Society Supplemental Volume* 9, 162–171.

Transformational Grammar: Evolution

S J Harlow

© 2006 Elsevier Ltd. All rights reserved.

This article is reproduced from the previous edition, volume 9, pp. 4650–4664, © 1994, Elsevier Ltd.

Transformational grammar was developed in the mid 1950s by Noam Chomsky. Over the next two decades it became the dominant paradigm in syntactic theory and description and its descendant, government binding theory is still one of the most influential

current theories. Transformational grammar forms a wide-ranging theory, whose central tenets are the use of hypothetico-deductive methodology to construct formal models of certain aspects of human linguistic capabilities. Such models are called 'grammars' in the theory and are taken to be an encoding, in some form, of the native speaker's linguistic knowledge (or 'competence'). Much of the work of transformational grammar has consisted of constructing models of (fragments of) individual languages, but equally important has been the task of exploring and defining

the properties which are required by such grammars to provide accurate and revealing accounts of the linguistic data under consideration. Such general properties are taken to form the content of 'linguistic theory' (also termed 'universal grammar') and therefore do not need to be stated in individual linguistic descriptions. While a grammar for an individual language is a representation of the linguistic knowledge of the native speaker of that language, linguistic theory represents the properties (possibly very abstract in nature) which constrain all languages and thus defines the notion 'possible human language.' The content of universal grammar is taken to be a characterization of the human language acquisition device – those antecedent conditions that make language acquisition possible and which constrain the learning space available to the child acquiring a native language. Transformational grammar has undergone significant evolution since its initial development and one of the slightly ironic consequences of the pursuit of the goals listed above is that transformations themselves have come increasingly to play a less and less significant role in the theory. This article traces that development.

Early Transformational Grammar

Transformations, from which the theory derives its name, were developed during the 1950s by Zellig Harris and Noam Chomsky. The central idea underlying the concept of transformation was that of capturing systematic relationships between different syntactic representations. However, because of fundamental differences in the metalinguistic goals of Harris and Chomsky, the role played by transformations in the work of the two and the consequences of their introduction also were radically different.

Harris on Transformations

For Harris, transformations offered a way of allowing structural linguistics to transcend the limits of the sentence and offer a method of approaching discourse-level structures. For Harris, transformations are a device that extends the descriptive linguistic techniques of segmentation and classification to texts larger than a single sentence. They allow the linguist to establish equivalences between sequences of words in a text which would otherwise resist analysis. For example, suppose that a text contains the sentences:

- (1) He played the cello.
- (2) The cello was played by Casals.

but no sentences of the form:

- (3) Casals played the cello.
- (4) The cello was played by him.

In other words, there are no environments which show that *Casals* and *he/him* are grammatically equivalent (i.e., belong to the same grammatical category).

If, however, the text contains active/passive sentence pairs:

- (5) He plays the guitar.
- (6) The guitar was played by him.

an equivalence can be established between the two sentence types, as follows:

- (7) $N_x V N_y \leftrightarrow N_y \text{ was V ed by } N_x$

Assuming that *He*, *him*, *Casals*, and *the guitar* belong to the category N , and that *played* is a member of the category V , (7) captures the relationship between the two; and between all other pairs of sentences possessing the same structure. With the equivalence given in (7) above, 'we can show that all the environments of *Casals* are equivalent to all those of *he*; and this in turn can make other equivalences discoverable textually' (Harris 1952: 129).

Chomsky on Transformations

For Chomsky, on the other hand, transformations formed part of a program directed at characterizing the nature and properties of the human language faculty. Chomsky states:

The development of these ideas that I would like to report on briefly, however, follows a somewhat different course. It begins by questioning the adequacy of a certain view of the nature of linguistic theory that has dominated much recent work, and it attempts to reformulate the goals of linguistic theory in such a way that questions of a rather different nature are raised. And finally, it attempts to show that the concept of grammatical transformation, in something like Harris' sense, but with certain differences, is essential to answering these questions (Chomsky, 1964b: 212).

One major difference between Chomsky's and Harris's applications of transformations was the nature of the representations between which a transformational mapping held. For Harris, the kinds of relationships captured by transformations were essentially those which in terminology developed later were called 'surface structure' relationships (Chomsky 1965a). That is to say, the mappings defined by transformations were limited to those which could be expressed in terms of the categories and linear order of the actually occurring words/morphemes in a sentence. For Chomsky, no such restriction was applicable. This is apparent in one of Chomsky's earliest and most well-known transformational analyses: that of the English auxiliary system. The abstractness permitted by Chomsky's conception

of transformations allowed him to propose an elegant and simple analysis of the auxiliary system which, in the words of Newmeyer (1980: 24), “[...] probably did more to win supporters for Chomsky than all of his metatheoretical statements about discovery and evaluation procedures.” In *Syntactic structures*, the book in which transformational grammar first became accessible to the linguistic community, Chomsky states:

The study of these ‘auxiliary verbs’ turns out to be quite crucial in the development of English grammar. We shall see that their behaviour is very regular and quite simply describable when observed from a point of view that is quite different to that developed above. Though it appears to be quite complex if we attempt to incorporate these phrases directly into a [S, F] grammar [i.e., a phrase structure grammar] (Chomsky 1957: 38).

and proposes the following phrase structure rules:

- (i) $Verb \rightarrow Aux \rightarrow V$ [28]
- (ii) $V \rightarrow \textit{hit, take, walk, read, etc.}$
- (iii) $Aux \rightarrow C (M) (\textit{have} + \textit{en}) (\textit{be} + \textit{ing}) (\textit{be} + \textit{en})$
- (iv) $M \rightarrow \textit{will, can, may, shall, must}$
- (i) $C \rightarrow \left\{ \begin{array}{l} S \text{ in the context } NP_{sing} \\ \emptyset \text{ in the context } NP_{pl} \\ \textit{past} [29] \end{array} \right\}$

Rule [28iii] states that an *Aux* phrase must contain as its initial element the tense/agreement morpheme *C*, and that this initial element can be followed by zero or more of *M*, *have + en*, *be + ing*, *be + en* in that order. *M* can then take any of the forms in [28iv]. Chomsky (1975: 232–33) observes, however, that:

the sequence of morphemes which results from an application of [28] is not in the correct order. Thus to complete the statement we give the following rule [which later became known as ‘affix hopping’]:

- (ii) Let *Af* stand for any of the affixes *past*, *S*, Φ , *en*, *ing*. [29]

Let *v* stand for any *M* or *V*, or *have* or *be* (i.e., for any nonaffix in the phrase *Verb*).

Then:

$Af + v \rightarrow v + Af \#$

where $\#$ is interpreted as word boundary.

- (iii) Replace $+$ by $\#$ except in the context $v \rightarrow Af$.
Insert $\#$ initially and finally.

When the items *S*, *have + en* and *be + ing* in rule [28iii] are selected to give the string:

- (8) $\textit{the} + \textit{man} + S + \textit{have} + \textit{en} + \textit{be} + \textit{ing} + \textit{read}$
 $+ \textit{the} + \textit{book}$

and rule [29ii] is applied three times (first to $C + \textit{have}$, then to $\textit{en} + \textit{be}$, and finally to $\textit{ing} + \textit{read}$), followed by

the application of rule [29iii], the following is derived:

- (9) $\textit{the} \# \textit{man} \# \textit{have} + S \# \textit{be} + \textit{en} \# \textit{read} + \textit{ing} \# \textit{the} \#$
 \textit{book}

Chomsky’s analysis thus depends upon an abstract structure in which the affixes of the auxiliary system do not appear in their ‘surface’ order. The question of what kind of syntactic representations could legitimately constitute the input and output of a transformation came to be one of the most significant issues in the development of transformational theory.

In Chomsky’s interpretation of the notion, a transformation is a mapping which converts a sentence (= a string of words) with an associated constituent structure analysis (= a ‘phrase marker’), defined by a set of phrase-structure rules, into a new sentence with a ‘derived constituent structure.’ A transformation consists of a ‘structural analysis’ (often also called a ‘structural description,’ and abbreviated to *SD*), which specifies the sequence of categories into which a phrase marker must be analyzable for the transformation to apply to it, and a ‘structural change,’ which specifies the changes effected on the input phrase marker by the transformation. For example, the passive transformation is given by Chomsky (1957: 112) in the following form:

- (10) Structural analysis:
 $NP - Aux - V - NP$
Structural change:
 $X_1 - X_2 - X_3 - X_4 \rightarrow X_4 - X_2 + \textit{be} + \textit{en} - X_3$
 $- \textit{by} + X_1$

(An alternative notation, which merges the structural analysis and structural change into a single statement of the form:

- $NP_1 - Aux - V - NP_2$
 $\rightarrow NP_2 - Aux + \textit{be} + \textit{en} - V - \textit{by} + NP_1$

was also commonly used in later transformational work.)

The structural analysis specifies that for the transformation to apply to a string, the phrase marker associated with that string must be analyzable into a sequence of $NP - Aux - V - NP$. In the structural change, the variables $X_1 - X_4$ identify the four categories listed in the structural analysis and the string to the right of the arrow defines the mapping effected by the transformation: the relative positions of the two NPs are interchanged, the morpheme *by* is adjoined to the left of X_1 in its new position and the morphemes *be* and *en* are adjoined to the left of X_2 (= *Aux*). Each transformation is thus the composition of a number of more elementary operations, which include

deletion, substitution of one element for another, adjunction, and permutation. For example, Chomsky (1965b: 120) specifies the operation of the passive transformation in terms of the following 'elementary transformation,' which details the operations performed by the rule:

- (11) $tp(Y_1; Y_1, \dots, Y_4) = Y_4$
 [substitute Y_4 for Y_1]
 $tp(Y_1, Y_2; Y_2, Y_3, Y_4) = Y_2 \wedge be \wedge en$
 [adjoin *be* and *en* to the right of Y_2]
 $tp(Y_1, Y_2, Y_3; Y_3, Y_4) = Y_3$
 [leave Y_3 unchanged (identity)]
 $tp(Y_1, Y_2, Y_3, Y_4; Y_4) = be \wedge Y_1$
 [adjoin *be* to left of Y_1]
 $tp(Y_1, \dots, Y_n; Y_n, \dots, Y_r) = Y_n$ for all $n \leq r \neq 4$
 [identity]

This notation identifies each term in the structural description of the transformation by successively splitting the structural description into two substrings (indicated in 11 by the semicolon), one of which terminates with that item and the other of which begins with it. For each such item, the elementary transformation specifies what operation is performed on it.

As stated above, Chomsky's motivation for the introduction of transformations was radically different to that of Harris. Chomsky had established a hierarchy of language types, which could be used to model the properties of human languages. Each of the languages on the Chomsky hierarchy is defined in terms of the possible sequences of symbols it admits. Chomsky presented a proof that English possesses properties which mean that it cannot be analyzed in terms of the most restricted type of language on the hierarchy: 'finite state languages.' He also argued that any attempt to describe the syntactic properties of English in terms of the kind of devices available for characterizing the next weakest language type on the hierarchy ('phrase structure languages') leads to the loss of generalizations and greater complexity. The simple and elegant analysis of the English auxiliary system or of the passive outlined above, for instance, which require the capacity to rearrange the order of elements in a string cannot be accomplished within the limits of these more restricted grammatical systems. Chomsky concludes (1957: 44) 'By further study of the limitations of phrase structure grammars with respect to English we can show quite conclusively that these grammars will be so hopelessly complex that they will be without interest unless we incorporate such rules [i.e., transformations].' (It is, however, important to note that Chomsky did not *prove* that English could not be analyzed by phrase structure

grammars at all. Rather, he argued that any such analysis would be unacceptably complex and unrevealing. Indeed, later developments in phrase structure grammars in the 1980s demonstrated that even this assertion is too strong.)

A grammar for Chomsky is a device which defines the language under investigation via a system of rules. In the earliest stages of transformational grammar (e.g., *Syntactic structures*), the grammar consisted of three sets of rules:

- (12) Phrase structure rules
 Transformations
 Morphophonemic rules

As seen above, the phrase structure rules defined the constituent structures which form the input to the transformational component. The morphophonemic rules are responsible for ensuring, for example, that *take + en* is realized as *took*.

Subtypes of Transformations

Transformations in the early period fell into two categories, 'singular' transformations, whose domain consisted of a single sentence, and 'generalized' (or 'double-based') transformations, with a structural description containing reference to more than one sentence. All the examples mentioned so far have been of singular transformations. An example of a generalized transformation is the rule for coordination (Chomsky 1957: 113):

- (13) Structural analysis: of S_1 : $Z - X - W$
 of S_2 : $Z - X - W$
 where X is a minimal element (e.g., *NP*,
 VP, etc.)
 and Z, W are segments of terminal strings.
 Structural change:
 $(X_1 - X_2 - X_3; X_4 - X_5 - X_6) \rightarrow (X_1 -$
 $X_2 + and + X_5 - X_6)$

For example:

- (Yesterday - John - left home; Yesterday - Sam -
 left home)
 \rightarrow (Yesterday - John + and + Sam - left home),
 where *John* and *Sam* are both instances of *NP*.

A further major class of generalized transformations were those that embedded one sentence in another. Chomsky (1964b) lists 19 such transformations, of which the following are representative examples:

- (14) Nominalization
 Structural analysis: S_1 : $T, it + \phi, VP$
 S_2 : NP, C, VP_1
 Structural change:
 $X_4 + S$ or ϕ replaces X_1 ; *ing* + X_6
 replaces X_2

For example:

$$\left\{ \begin{array}{lll} \text{T} & -\text{it} + \text{C} & - \text{be} + \text{a} + \text{great} + \text{surprise} \\ \text{X1} & \text{X2} & \text{X3} \\ \text{John} & - \text{C} & - \text{prove} + \text{the} + \text{theorem} \\ \text{X4} & \text{X5} & \text{X6} \end{array} \right\}$$

→ John + S - ing + prove + the + theorem - C -
be + a + great + surprise

giving 'John's proving the theorem was/is a great surprise.'

(15) Complement

Structural analysis: $S_1: X, V_T, \text{Comp}, NP$
 $S_2: NP, \text{Aux}, \text{be}, \text{Pred}$
 Condition: X_2 is a member of the
 class of verbs
 containing *consider*,
believe, ...
 Structural change: X_5 replaces X_4 ; X_8
 replaces X_3

For example:

$$\left\{ \begin{array}{llll} \text{They} + \text{C} & - \text{consider} & - \text{Comp} & - \text{the} + \text{assistant} \\ \text{X1} & \text{X2} & \text{X3} & \text{X4} \\ \text{the} + \text{assistant} & - \text{C} & - \text{be} & - \text{qualified} \\ \text{X5} & \text{X6} & \text{X7} & \text{X8} \end{array} \right\}$$

→ They + C - consider - qualified - the + assistant

Following the application of an obligatory transformation which permutes the last two factors in the derived phrase marker, this gives *They consider the assistant qualified*.

This class of generalized transformation exemplified by (15) formed the locus of recursion in early transformational grammar, and was thus the mechanism by which grammars make available an infinite set of sentences from finite resources.

Rule Ordering

Transformations, in principle, could be ordered arbitrarily relative to one another (i.e., they were extrinsically, rather than intrinsically ordered). For example, the derivation of an interrogative sentence such as *Did John eat an apple* involved the application of the (optional) question-formation transformation:

(16) Structural analysis: $\left\{ \begin{array}{l} NP - C - V \dots \\ NP - C + M - \dots \\ NP - C + \text{have} - \dots \\ NP - C + \text{be} - \dots \end{array} \right\}$

Structural change: $X_1 - X_2 - X_3$
 → $X_2 - X_1 - X_3$

In other words, the tense formative *C* moves to the front of the sentence. The resulting derived P-marker requires the obligatory application of the word-boundary transformation [29iii] and the *do*-insertion

(17) *do*-insertion

Structural analysis: $\# - Af$
 Structural change: $X_1 - X_2 \rightarrow X_1 - do + X_2$

(Adjoin the word *do* to the left of any affix immediately preceded by a word boundary.)

The successive stages of the derivation would look like this:

Phrase structure: John - C - eat - the apple
 Question-formation: C - John - eat - the apple
 Word-boundary: #C#John#eat# the apple
 Do-insertion: #do + C#John#eat# the apple

(*do* + *C* is then converted to *did* by a morphophonological rule).

In order to correctly derive the sentence *Who ate the apple*, it was necessary to stipulate that T_w , the rule which moves an interrogative element such as *who* to sentence initial position, must apply *after* the question-formation transformation:

Phrase structure rules: John - C - eat - the apple
 Question-formation: C - John - eat - the apple
 T_w : Who - C - eat - the apple
 Word-boundary: #Who#C eat# the apple

Do-insertion cannot apply to this structure, so affix hopping [29ii] applies obligatorily instead:

Affix Hopping: #Who# eat + C# the apple

Applying the two transformations in the reverse order would have the following effect:

Phrase structure rules: John - C - eat - the apple
 T_w : Who - C - eat - the apple
 Question-formation: C - who - eat - the apple
 Word-boundary: #C # who# eat# the apple
 Do-insertion: #do + C# who# eat# the apple

This would give the ungrammatical **Did/does who eat the apple*. The only way in this analysis to avoid the generation of this ungrammatical string is to stipulate the ordering which gives the correct results.

In addition, a stipulation was required as to whether the application of a given transformation was either obligatory or optional. In the examples given above, the passive transformation is an instance of an optional rule, 'affix hopping,' [29ii] an instance of an obligatory rule - it must apply to any structure which meets the terms of its structural analysis. A sentence that is generated solely by the application of obligatory rules was termed a 'kernel' sentence. These corresponded to the class of simple, active, declarative sentences. Associated with the transformational derivation of each sentence was a representation called a 'T-marker,' which recorded which transformations had applied and in which order.

Transformations and Mental Processes

Although transformational grammar is intended as a model of the linguistic knowledge of the native speaker, it was not Chomsky's claim that the components of the theory were a model of the way the construction or comprehension of linguistic utterances is computed in the brain. Nonetheless, a number of experiments carried out in the 1950s and 1960s seemed to show that the transformational grammars developed for English during the early period did indeed make accurate predictions if taken as models of mental processing. In brief, these experiments seemed to show that there was a correlation between the number of transformations involved in the derivation of a sentence and the processing complexity associated with the sentence, measured in terms of some behavioral index such as response time – the greater the number of transformations involved, the greater the processing difficulty. For example, passive sentences take longer to process than their active counterparts and negative sentences take longer to process than their affirmative counterparts.

However, subsequent studies and a more mature assessment of the results of these experiments have shown that those experiments which seemed to show a positive correlation between transformational complexity and processing load did so because of a failure to control for other variables. It is well attested that processing load increases with sentence length. In the case of passive sentences, for example, *Albert was eaten by a lion* is two words longer than *A lion ate Albert*. Once such factors as this were controlled for, it became apparent that the derivational complexity hypothesis could not be maintained. It is universally agreed that the relationship between the form of a transformational grammar and the manner in which it is implemented in the brain is much more abstract and indirect than these early experiments suggested. It is therefore important to note that when linguists talk of a grammar 'generating' a set of sentences, there is no implication in the use of the term that the grammar is a model of human sentence production (see Carroll and Bever 1976).

The 'Standard Theory'

As Chomsky comments (1975: 23), "The theory of transformations developed in *The Logical Structure of Linguistic Theory* was far too unrestricted. It seems that there are much narrower conditions that determine the applicability of transformations and the kinds of mapping they can perform." The subsequent development of transformational grammar can be seen as a progressive transfer of many of the functions

originally assigned to transformations to other components of the grammar.

There are five major areas in which this pattern of development can be seen:

1. Rule interaction;
2. Recursion;
3. Derived constituent structure;
4. The relationship of syntax to semantic interpretation;
5. The degree of abstractness permitted for underlying representations.

Rule interaction had presented a problem in the early theory of transformations because it was not clear what kind of ordering relationships could exist between singular and generalized transformations. It was not clear whether, for example, singular transformations could apply to an incomplete sentence structure *before* a generalized transformation embedded material in it. Various suggestions were made concerning the 'traffic rule' problem, but the most significant contribution was that of Fillmore (1963) who observed that it did not appear that there were any cases in which a singular transformation had to apply to a structure before a generalized transformation embedded material in it. Second, it did not appear that it was necessary in practice to stipulate any ordering between generalized transformations, and third, there were attested cases where a singular transformation could apply to a sentence before it was embedded in a larger structure (passive, for example, in sentences such as *The director believed the keeper to have been eaten by one of the lions*). Fillmore's conclusion from these facts was that singular transformations should apply first to (what would be) the most deeply embedded sentence. This sentence would then be embedded in the matrix sentence by a generalized transformation. This sentence would then become the domain of the singular transformations again, and the process would be repeated as often as necessary.

Since transformations are mappings from phrase markers to phrase markers, it is essential that there be some well-defined way of specifying the constituent structure that results from the application of a transformation. The passive, for example, requires that the word *by* be adjoined to the left of the NP which has been moved into final position. It is clear that the resulting constituent should be a prepositional phrase, and that *by* itself should be a preposition. This information is not derivable, however, from the version of the transformation given above. Some linguists resolved this issue by specifying the details of the derived constituent structure in the structural change of the transformation itself:

$$(18) NP_1 - Aux - V - NP_2 \rightarrow NP_2 - Aux + be + en \\ - V - [_{PP}[_P by] + NP_1]$$

This increasingly came to be considered an undesirable enrichment of the theory and alternatives were sought.

The problem of derived constituent structure also arose in the context of the application of generalized transformations. As can be seen from the complement example above (15), the transformation not only embeds material from one clause into another, but also deletes the subject of the embedded clause. The question then arises as to what the category of the embedded constituent is. The idea rapidly developed (Katz and Postal 1964) that clauses into which a sentence was to be embedded should contain a 'dummy' node, defined by the phrase structure component of the grammar, which would provide a host (and a category) for the embedded material.

These issues were addressed in the 'standard theory,' the name given to the revised version of transformational grammar developed by Chomsky in his book *Aspects of the theory of syntax* (1965a).

Rule Interaction: the Transformational Cycle

In *Aspects* the issues listed above received a radical answer, which reduced the role which transformations played in the theory. Most dramatically, the class of generalized transformations was eliminated entirely and their role as the locus of recursion was allocated to the phrase structure component. The issue of rule interaction was resolved by combining Fillmore's proposals with this innovation. Their joint effect is to require that transformations apply first to the most deeply embedded sentential structure, and, only when all possibilities have been exhausted, is the next 'higher' sentential level considered as a transformational domain. Graphically, the situation is illustrated in the schematic tree in Figure 1.

All transformations apply first to S_2 , then to S_1 , and only then to S_0 . This principle of transformational rule application is known as the ('transformational') cycle.

By way of specific illustration, take the analyses of complementation by Rosenbaum (1967) whose thesis has proved a reference point for most subsequent work. Rosenbaum's work is concerned with the analysis of sentences such as the following:

- (19) Sam prefers to stay here.
- (20) I believe Bill to have been convinced by John.
- (21) John is believed by me to have convinced Bill.

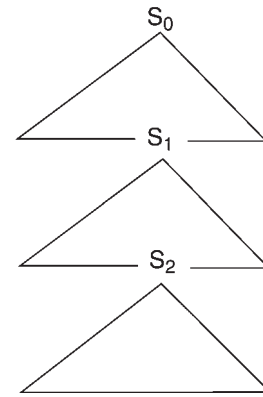


Figure 1

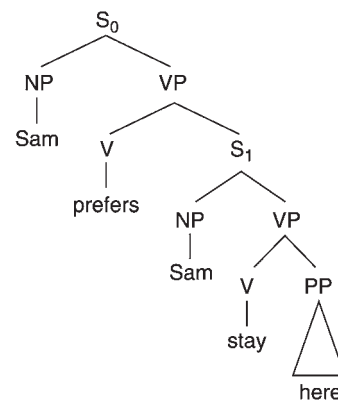


Figure 2

Each of these sentences would earlier have involved a generalized transformation. In the *Aspects* framework, the structures underlying them are provided by the following phrase structure rule:

$$(22) VP \rightarrow V (NP) S$$

This rule preempts the role of generalized transformations by introducing the symbol S on its right-hand side. The structure associated with (19) by Rosenbaum's phrase structure rules (suppressing some details) is given in Figure 2.

On the lowest cyclic domain (the structure dominated by S_1) no rules apply. When the next highest cyclic domain (S_0) is reached, a 'complementizer placement' transformation is responsible for transforming this basic structure into:

Second Cycle: Sam prefers [_{VP} to stay here]]

This structure meets the structural description of what Rosenbaum termed the 'identity erasure transformation' (now generally known as 'equi NP deletion,' or simply 'Equi'):

(23) Equi NP Deletion

Structural Description:

W	(NP)	X	+D	NP	Y	(NP)	Z
1	2	3	4	5	6	7	8

(Where +D identifies the complementizer
for – to.)

Structural Change: 5 is erased (deleted) by 2

The effect of the application of this rule on *Sam prefers [for Sam [VP to stay here]]* is to delete the second occurrence of the NP *Sam*:

Second Cycle: Sam prefers [for [VP to stay here]]

which finally gives *Sam prefers [VP to stay here]]*, through the application of a rule of ‘complementizer deletion.’

For (20) and (21), Rosenbaum’s analysis would provide the derivations in (20a) and (21a). (Note the use of the substructure [MAN by + P] in these derivations to provide a trigger for the passive transformation and a landing-site for the NP moved from subject position by the passive.)

(20a) First Cycle:	I believe [NP it [S [NP John] [VP have + en convince Bill] [MAN by + P]]]
Passive:	I believe [NP it [S [NP Bill] [VP have + en <i>be</i> + en convince [MAN by + <i>John</i>]]]]
Affix Hopping:	I believe [NP it [S [NP Bill] [VP have <i>be</i> + en convince + en] [MAN by + John]]]
Second Cycle:	
Complementizer Placement:	I believe [NP it [S for [NP Bill] [VP to have <i>be</i> + en convince + en] [MAN by + John]]]
Extraposition:	I believe [NP it] [S for [NP Bill] [VP to have <i>be</i> + en convince + en] [MAN by + John]]
Raising to object:	I believe [NP <i>Bill</i>] [S for [VP to have <i>be</i> + en convince + en] [MAN by + John]]
Complementizer Deletion:	I believe [NP Bill] [S [VP to have <i>be</i> + en convince + en] [MAN by + John]]
Post Cycle:	
Morphophonemic Rules:	I believe [NP Bill] [S [VP to have <i>been convinced</i>] [MAN by + John]]

(‘Extraposition’ is a transformation that takes a structure of the form [NP it S] and moves the S rightwards out of the NP.)

(21a) First Cycle:

I believe [NP it [S [NP John] [NP have + en convince Bill]]] [MAN by + P]

Affix Hopping:

I believe [NP it [S [NP John] [VP have convince + en Bill]]] [MAN by + P]

Second Cycle:

Complementizer Placement:

I believe [NP it [S for [NP John] [VP to have convince + en Bill]]] [MAN by + P]

Passive:

[NP it [S for [NP John] [VP to have convince + en Bill]]] *be* + en believe [MAN by + I]

Extraposition:

it *be* + en believe [MAN by + I] [S for [NP John] [VP to have convince + en Bill]]]

Raising:

John *be* + en believe [MAN by + I] [S for [VP to have convince + en Bill]]]

Affix Hopping:

John *be* *believe* + en [MAN by + I] [S for [VP to have convince + en Bill]]]

Complementizer Deletion:

John *be* *believe* + en [MAN by + I] [S [VP to have convince + en Bill]]]

Post Cycle:

Morphophonemic Rules: *John* *is* *believed* [MAN by + *me*] [S [VP to have convinced Bill]]]

In (20a) the passive transformation applies on the first cycle (inside the subordinate clause), whereas in (21a) it applies on the second cycle (in the main clause). In both derivations affix hopping applies twice, once on the first cycle and once on the second. Both these sentences also exemplify the rule of raising (called ‘pronoun replacement’ by Rosenbaum), which moves the subject of a nonfinite subordinate clause (here *John*) into the next higher clause on the second cycle. The existence of this rule was to become a major issue in the subsequent development of transformational grammar.

The Organization of the Grammar

With the adoption of the transformational cycle, many of the problems of rule interaction were satisfactorily resolved and, by the time of the publication of *Aspects of the theory of syntax*, the organization of the grammar took the following form (shown in Figure 3).

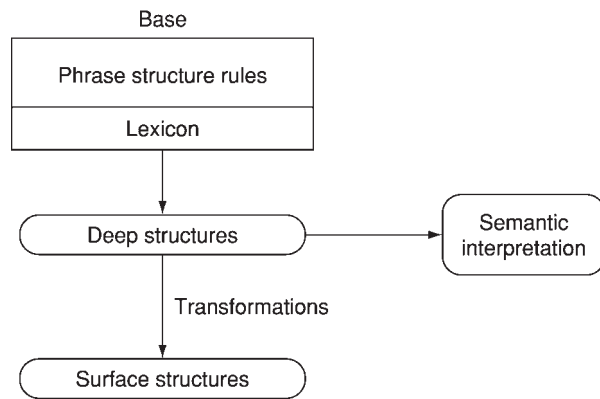


Figure 3

The ‘lexicon’ (formerly part of the phrase structure component) and Phrase Structure Rules were split into separate subcomponents of the ‘base component’ of the grammar and jointly determined a level of syntactic representation called ‘deep structure.’ This, on the one hand, provided the input to rules of semantic interpretation and, on the other, the input to the transformational component of the grammar, which mapped deep structure representations on to surface structure representations in the manner illustrated in (20a) and (21a) above.

Generative Semantics

The proposal that deep structure formed the level at which both lexical insertion and semantic interpretation took place came under strong challenge during the late 1960s. The branch of transformational grammar which became known as generative semantics argued that this conception of the organization of the grammar led to a loss of generalizations, a situation which could be remedied if underlying representations were allowed to be more abstract than Chomsky’s deep structures. This aspect of generative semantics was driven by the Katz–Postal hypothesis (Katz and Postal 1964) concerning the relationship between syntax and semantics. This proposed that all the information necessary for semantic interpretation is represented in the underlying structure of a sentence. Corollaries of this proposal are that synonymous sentences have a single underlying structure and that ambiguous sentences have more than one underlying structure. It also implies that transformations do not change semantic interpretation.

For many sentence pairs, such as actives and passives, this proposal seems eminently reasonable. The assumption that *John saw Bill* and *Bill was seen by John* have the same deep structure and acquire their

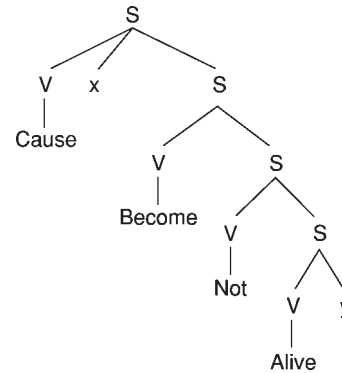


Figure 4

different surface forms through the application of transformations provides a basis for accounting for the fact that the two sentences have the same truth conditions. However, this logic, when applied to *Bill is dead* and *Bill is not alive* leads to the conclusion that they too must share a single underlying structure. In this case, however, such an identity in underlying structure must exist *before* lexical insertion, contradicting Chomsky’s hypothesis about deep structure. This led McCawley (1968) to propose that the underlying structure of a sentence of the form *x killed y* should receive the analysis in Figure 4 (note that McCawley is assuming here that in its underlying structure English is verb-initial; see McCawley 1970 for details).

By the repeated cyclic application of an optional transformation of ‘predicate raising,’ the verbs in this structure could be adjoined to one another to give the structure in Figure 5.

McCawley proposed that ‘if lexical insertion did not apply until after that transformation, then the “dictionary entry” for *kill* could be expressed as a transformation which replaces the subtree at the left [of the above tree] by *kill*’ (McCawley 1968: 73). If lexical insertion is constrained to apply before all transformations, as in the *Aspects* approach, the semantic relationship between *x killed y* and *x caused y to become not alive* could not be expressed using transformations.

As an example of ambiguity prompting the postulation of different underlying structures, where purely syntactic evidence might not warrant them, take Lakoff’s (1970) argument that the ambiguity of the sentence (24), with its two paraphrases (25) and (26):

(24) I don’t beat my wife because I like her.

(25) It is because I like her that I don’t beat my wife.

(26) It is not because I like her that I beat my wife.

could be accounted for by the assumption that the two readings correspond to different underlying structures. For (25), Lakoff proposed the structure in Figure 6.

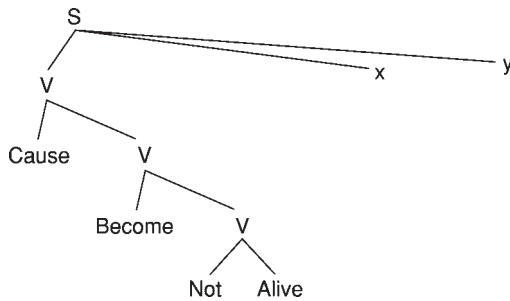


Figure 5

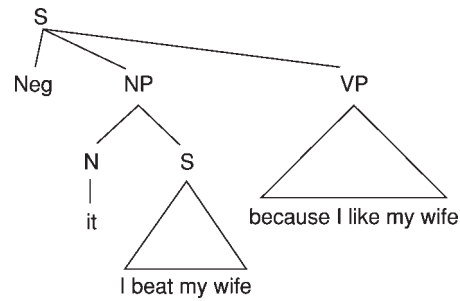


Figure 7

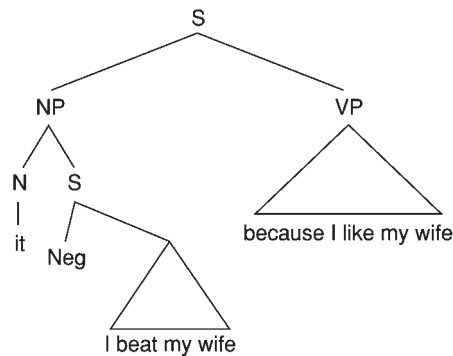


Figure 6

For (26), he proposed the structure in Figure 7.

Transformations called 'it-deletion' and 'neg lowering' apply to these structures to give (24).

A further modification of standard transformational grammar made within generative semantics concerned the nature of the relationship which held between the various levels of a derivation. In standard transformational grammar, transformations are in effect relations between adjacent stages in a derivation. There was no possibility of a transformational 'looking forward' to a future stage in the derivation, for example. Generative semanticists, on the other hand, argued that such extensions were necessary. Lakoff (1971) observed that sentences like *Many men read few books* and *Few books are read by many men* seem to differ in their interpretations. The first is paraphrasable as *There are many men who read few books* and the second as *There are few books that are read by many men*. Lakoff proposed that these two interpretations are derived from different underlying structures. The first has the underlying structure in Figure 8, the second that in Figure 9.

The surface structures are derived by a transformation of 'quantifier lowering.' A problem remains, however, in ensuring that structure in Figure 8 must undergo passive and the structure in Figure 9 must not. Lakoff's solution was to propose that passive and

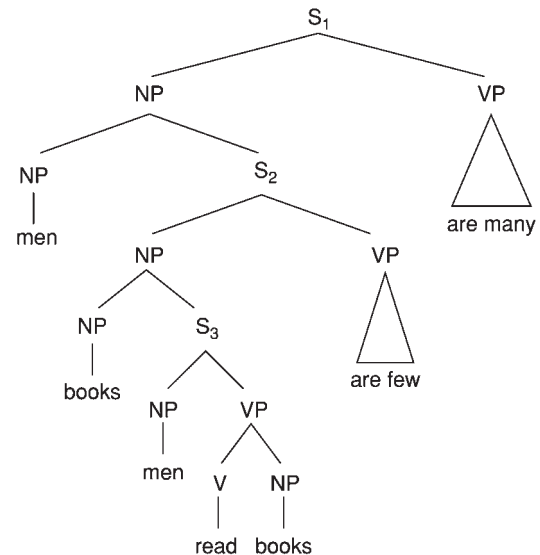


Figure 8

quantifier lowering could freely apply to both structures, but that a 'global derivational constraint' checked the relationship between the underlying and surface forms, requiring that the quantifier which was highest in underlying structure must be the leftmost one in surface structure. Global constraints allow the application of a transformation to have access to any level of representation in the derivation, not merely the information coded in its structural description and structural change.

In essence, what generative semantics did was to pursue the use of transformations to express linguistic regularities and subregularities to its logical conclusion. It assumed that transformations were the *only* way that such regularities should be captured, incorporating into syntactic description phenomena from semantics, pragmatics, and even social interaction. Ultimately, however, this central tenet of generative semantics was rejected by the majority of linguists working in the transformational framework. (A more comprehensive discussion of generative semantics and the factors both linguistic and nonlinguistic

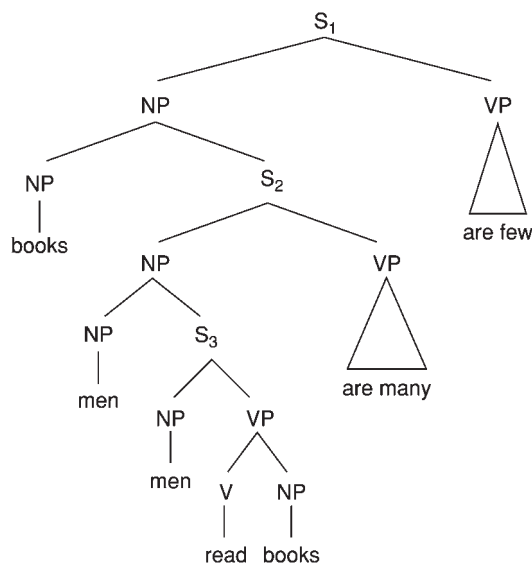


Figure 9

involved in its rise and fall can be found in Newmeyer 1980: ch. 4 and 5.)

The Extended Standard Theory

Subsequent developments in transformational grammar were motivated by a desire to constrain the expressive power of the transformational component. These took two forms. One was to seek to impose restrictions on transformations themselves, with the goal of limiting their expressive power; the other was to reallocate the responsibility for certain classes of phenomena to other components of the grammar, in particular the lexicon, the phrase structure component, and the semantic component. The name given to the version of the theory incorporating these revisions, which developed during the 1970s, was the 'extended standard theory.'

The work of Peters and Ritchie (1973) too undoubtedly encouraged a climate in which constraining the power of transformations was seen as desirable. They provided a proof that transformational grammars of the *Aspects* type had the weak generative capacity of the most powerful device on the Chomsky hierarchy, unrestricted rewriting systems (Turing machines). This meant that human languages, as characterizable by transformational grammars, had no particular properties other than that they could be generated by some set of well-defined rules. *Aspects*-style transformational grammar did not place any interesting constraints on the class of possible human languages. In terms of language acquisition, this implies that the formal properties of transformational grammars provided no interesting restrictions on the search space accessible to the

learner. There was undoubted merit, in the light of these results, in any program which sought to constrain the power of existing transformational grammars.

The Lexicalist Hypothesis

'Nominalization,' referred to above, had been one of the earliest transformational rules proposed (see Lees 1966). Yet in 'Remarks on nominalization,' Chomsky (1970) developed the argument that at least some nominals which had previously been assumed to be transformationally derived were in fact simply basic lexical entries. His argument had two prongs. First, he demonstrated that the basis for the purported generalization was unsound and, second, that there were alternative ways of capturing the generalizations which remained.

In arguing against a transformational analysis of nominalization, Chomsky noted that what he termed 'derived nominals' (such as *refusal*, *easiness*, *revolution*, *belief*) do not occur in structures which are themselves derived by transformations. (The term 'derived nominal' does not imply any transformational derivation. Rather, the term is an allusion to the fact that such nominals form part of the derivational morphology of English.) So the sentences on the right of the arrows in (27) and (29) below are well-formed, but their corresponding derived nominals in (28) and (30) are not:

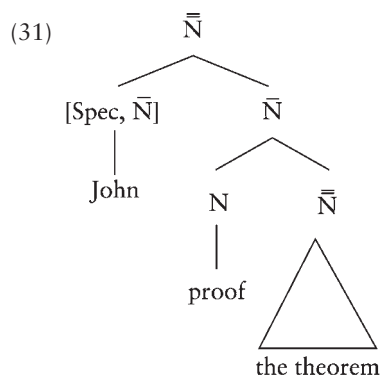
- (27) To please John is easy \Rightarrow John is easy to please
- (28) *John's easiness to please
- (29) For John to win the prize is certain \Rightarrow John is certain to win the prize
- (30) *John's certainty to win the prize

Furthermore, the semantic relationship which holds between a derived nominal and the item it is supposed to be derived from is frequently idiosyncratic. *John's deeds*, for example, are not the same as *things which John did*, nor is *John's ignorance* the same as *what John ignores*. In the common situation where there is no verb or adjective corresponding to a derived nominal (such as 'book,' 'poem'), the proponent of a transformational analysis is forced to postulate abstract underlying verbs such as *poetize* and to formulate mechanisms to ensure that these nonexistent items do not show up in surface structures (see, for example, Lakoff 1970).

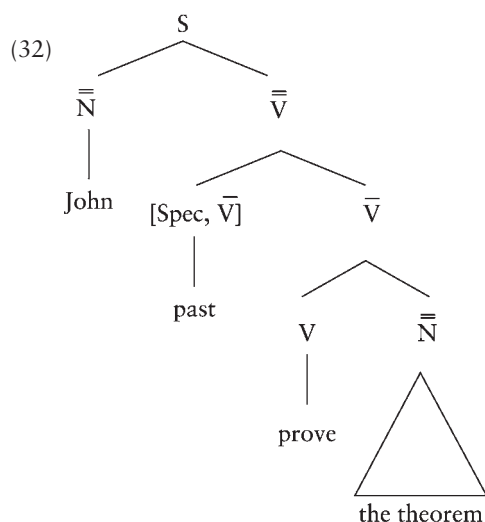
To capture the regularities that do occur, such as *John refused to leave/John's refusal to leave*, *The enemy destroyed the city/The enemy's destruction of the city* and *The city was destroyed by the enemy/The city's destruction by the enemy*, Chomsky proposed first that in cases where there exists a verb/derived

nominal pair, the lexical entry should be neutral with respect to grammatical category, but should specify what complements the pair require. In addition, he proposed an extension of the theory of phrase structure, known as 'X-bar theory,' to allow generalizations to be made across different linguistic categories.

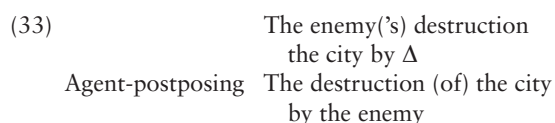
For example, sentences and NPs have a similar internal structure (cf. Chomsky, 1970: 211). In (31) *the theorem* is the direct object of the noun *proof* and *John* is its subject. (This structure also requires the later insertion of the preposition *of*.)



In the corresponding sentence, *the theorem* and *John* hold exactly the same grammatical relations to the verb *prove*.



Chomsky also proposed that cyclic transformations such as passive could apply not only within sentences, but also within NPs, and proposed factoring the passive transformation into two subrules: 'NP-preposing' and 'agent-postposing.' The derivation of *The city's destruction by the enemy* would look like this (with various details omitted):



NP-preposing The city('s) destruction by
the enemy

Note that the intermediate stage, resulting from agent-postposing, is also grammatical, provided that the NP's subject position is filled by a determiner. Exactly the same transformations apply in the case of the sentential passive *The city was destroyed by the enemy*, with the additional proviso that, for independent reasons, NP-preposing is obligatory.

Chomsky's argument, then, was that the generalizations captured by the nominalization transformation were spurious and, to the extent that genuine generalizations exist (such as the existence of active/passive pairs of nominalizations) there are alternative analyses available. The upshot of this was the proposal that transformations cannot change the syntactic category of an item. Thus, the rule of nominalization, as posited in early transformational grammar is not a possible transformation. This proposal was termed the 'lexicalist hypothesis,' because it assumes that any relationship of a crosscategorical kind is represented in the lexicon, and not by a transformational relationship.

Interpretive Semantics

The lexicalist hypothesis removes from the transformational component the power to alter syntactic categories and hence results in the reallocation of one class of linguistic phenomena to another component of the grammar. This is a tendency that is seen to be repeated with respect to other linguistic phenomena as the theory develops.

From the early days of transformational grammar, it had been assumed that the relationship that exists between an anaphoric pronoun and its antecedent, in sentences such as (34), where *Alice* and *she* can be interpreted as referring to the same individual, was to be captured by a transformation:

(34) Alice said that she should go.

The standard analysis of anaphoric pronouns (Lees and Klima 1969) assumed that they were derived from an underlying structure containing two identical occurrences of an NP. One of the NPs would be obligatorily converted into a pronoun by the pronominalization transformation:

(35) Pronominalization
 $X - NP_1 - Y - NP_2 Z \rightarrow X - NP - Y - NP_2 [+Pro] Z$

where $NP_1 = NP_2$ [and various conditions on the topological relations between the two NPs were satisfied].

Jacendoff (1972) argued instead for a treatment of pronominal anaphora which relied on rules of semantic

interpretation (an 'interpretive' approach to pronominalization). Jackendoff pointed out that the condition requiring identity between the two NPs (which must include not only syntactic identity but also identity of reference) is unexpected given that coreference is not generally a condition on the application of transformations; no movement rule, for example, is constrained to apply only to coreferential NPs.

Jackendoff's proposal assumed that anaphoric pronouns were base-generated in underlying structure and that rules of semantic interpretation established the necessary connection between them and their antecedents; no transformation of any kind was involved. This approach had a number of empirical consequences. It provided the basis for a solution to the problem that, in sentences containing quantified NPs such as *every runner*, the transformational treatment of pronominalization predicts incorrectly that *Every runner expects that she will win* should have the interpretation associated with *Every runner expects that every runner will win*.

Furthermore, there are cases of anaphora which do not involve pronouns. Such sentences as *I wanted to help Harry, but the fool wouldn't let me* allow a coreferential interpretation of *Harry* and *the fool*. In addition, the conditions on the configurations in which anaphora is possible are shared by pronominal anaphors and by 'epithets' like *the fool*. For example, neither permit an anaphoric interpretation in sentences like the following *He/the fool said Harry would leave*. It is implausible to envisage a transformation giving rise to epithets, but no problems arise in principle in assuming that epithets are assigned interpretations by the same semantic rules that interpret pronouns. Note also that Jackendoff's proposal is consonant with the lexicalist hypothesis in not allowing a transformation to radically alter the morphological shape of lexical material.

The net conclusion of these and other arguments is that pronominalization is not a transformation and that the phenomenon of pronominal anaphora falls within the domain of the semantic component of the grammar.

Similar arguments apply to two other long-established transformational analyses; those of reflexivization and equi-NP deletion. Reflexivization was assumed to work in complementary distribution with pronominalization; roughly, where the latter applied to identical NPs which were separated by a clause boundary, the former applied to identical NPs in the same clause. Thus, *Mary was proud of Mary* would give *Mary was proud of herself*. Note, however, that when a quantified NP is involved, the result is the same as in the case of pronominalization; *Every runner is proud of herself* does not have the same interpretation as *Every runner is proud of every*

runner. Equi-NP deletion, which derives *Mary wanted to win* from *Mary wanted Mary to win* exhibits the same phenomenon; *Every runner wanted to win* does not have the same interpretation as *Every runner wanted every runner to win*. As a consequence, proponents of the interpretive semantics position argued that neither reflexives nor the control phenomena handled by equi-NP deletion were the responsibility of the transformational component. Instead, reflexive pronouns should be basegenerated and interpreted by the semantic component. For 'equi' sentences it was proposed that the subject position of the nonfinite subordinate clause was filled by a basegenerated phonologically empty pronominal (PRO), also interpreted by the semantic component.

The result of these proposals was the removal from the transformational component of responsibility for a whole class of linguistic phenomena – those involving coreference – resulting in a more constrained theory of transformations and a superior account of the empirical data.

Constraints on Movement Rules

Structure Preservation The lexicalist hypothesis and interpretive semantics reduced the range of phenomena to be accounted for by the transformational component. For those transformations that remained, a major goal of the extended standard theory was to remove from the theory the capacity for individual transformational rules to contain detailed specifications of the conditions of their application, and to replace them with general conditions applying to all transformations, or to whole classes of transformations. This goal may be summarized as that of restricting the expressive power of transformations.

It was noted above that the precise manner in which the derived constituent structure resulting from the application of a transformation was determined was unclear and that various *ad hoc* measures were adopted, such as stipulating derived constituent structure in the structural change of the rule or introducing 'dummy' elements to act as the landing site of a moved constituent. This issue received a more highly-motivated solution in proposals by Emonds (1976). Emonds established three different classes of movement transformations:

(36) Root transformations

Those that crucially involved the root (roughly speaking the topmost) S node, and hence do not apply in subordinate clauses. Examples include Subject–Auxiliary Inversion (*Isn't it cold today*), Directional Adverb Preposing (*Down the street rolled the baby carriage*) and Topicalization (*Our daughters we are proud of*).

(37) Local transformations

Those that involve the reordering of two adjacent constituents, one of which is phrasal and the other nonphrasal. An example is Particle Movement (cf. *She picked up the book* and *She picked the book up*).

(38) Structure Preserving Transformations

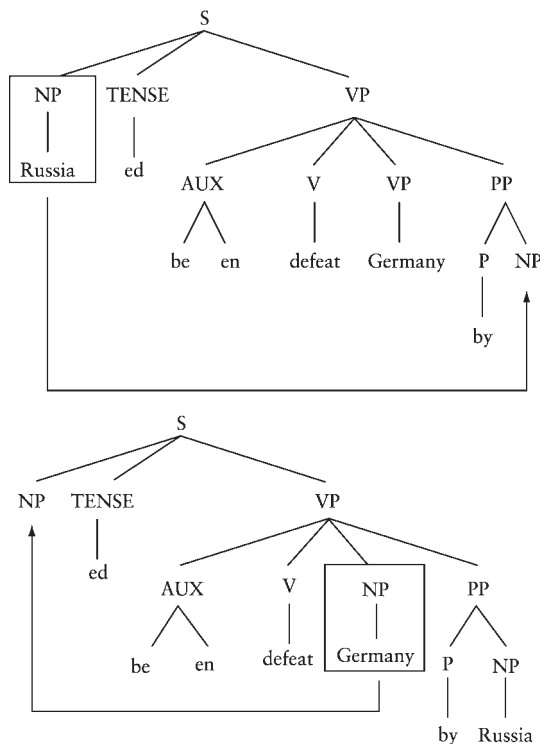
Those that involve the replacement of a node by one of the same syntactic category. Examples include Passive, Raising, There-insertion (*A book is on the table* \Rightarrow *There is a book on the table*) and Dative Movement (*gave a ticket to John* \Rightarrow *gave John a ticket*).

The phenomena handled by root and local transformations have been subject to reanalysis more recently, but the essence of the structure preserving hypothesis has been accepted by all subsequent work.

Emonds argued, like Chomsky, that the passive transformation is in fact a composition of two separate rules: one that postpones the agent NP into the position of *by*-phrase object; and one that preposes the object NP into the vacant subject NP position. This requires that the derivation of *Germany was defeated by Russia* looks like this:

(39)

Agent Postposing:



The structure-preserving hypothesis requires that the *by*-phrase be present in deep structure and that the NP within it be lexically empty (note that this implies that lexical insertion need not take place at deep structure). This analysis accounts for the fact

that the passive *by*-phrase is like any other PP with respect to its internal structure and transformational behaviour.

Emonds points out that the fact that PPs under VP are generally optional in English, when combined with the structure-preserving analysis of passive, accounts automatically for the existence of agentless passives (*Germany was defeated*), in contrast to early analyses which had postulated a rule of *by*-phrase deletion.

Emonds's proposals constituted a step in the direction of removing from the structural description and structural change of a transformation the necessity, or even the possibility, of specifying where a constituent could be moved to; e.g., if the structure involved is not a root S and the transformation does not involve two adjacent constituents, the only possibility is for the moved constituent to move to a location in which the phrase structure rules permit a constituent of that type to be generated.

Blind Application of Transformations The 1970s saw major steps toward the goal of replacing conditions on specific transformations with general conditions which, instead of forming part of individual grammars, were a component of universal grammar. Chomsky (1973) noted that, if the structural description of the passive transformation in English were taken to be essentially:

(41) X, NP, V, NP, Y

with no reference to conditions requiring the two NPs to be members of the same clause (as had been standardly assumed), the transformation would apply to the deep structure analysis of a sentence like *I believe the dog to be hungry*, namely [_S I believe [_S [_{NP} the dog] [_{VP} to be hungry]]] without the necessity for raising to move the NP *the dog* into the higher clause. Without any restriction on its operation, passive (and other transformations) would apply 'blindly' to any structural configuration meeting the structural description. Viewed in this way there is no problem in accounting for sentences such as *The dog is believed to be hungry by me*. Instead, there is a question as to why passive does not apply to *I believe the dog is hungry* to give **The dog is believed is hungry by me*. Chomsky's conclusion was that the failure of the passive to apply to the latter sentence is to be attributed to a general condition on transformational rules to the effect that no item may be extracted from a tensed (finite) clause. This constraint is phrased as follows:

(42) The Tensed-S Condition (TSC)

No rule can involve X, Y in the structure
 ... X ... [_a ... Y ...] ... [where *a* is a tensed sentence]

Chomsky's discussion of reciprocals (*each other*) illustrates the application of the TSC. Following Dougherty (1970), he assumes that *each* in sentences such as (43):

- (43) The men hated each other.

originates in deep structure as part of the subject NP as in (44):

- (44) The men each hated the others.

and is moved out of it by a transformation ('each' movement) to give the structure in (43). The TSC explains why this movement is not possible in sentences like *The men each expected* [_s *the others would win*] to give **The men expected* [_s *each other would win*], and why movement of *each* is possible in the nonfinite counterpart *The men each expected* [_s *the others to win*], giving *The men expected* [_s *each other to win*].

However, given the grammaticality of *The men expected* [_s *to shoot each other*], in which *each* appears in the object NP of the lower clause, the TSC does not account for the ungrammaticality of **The men expected* [_s *the soldier to shoot each other*]. To account for this paradigm, Chomsky proposed the specified subject condition:

- (45) Specified Subject Condition (SSC)
No rule can involve X, Y in the structure
... X ... [α ... Z ... – WYV* ...]. ...

where Z is the subject of WYV in α , α is a cyclic node (NP or S) and α contains a subject distinct from Z and not controlled by X.

'Control' is the relationship which holds between PRO and its antecedent in 'Equi' constructions, or between a 'trace' (see below) and its antecedent. Any subject meeting the additional criterion in (45) is a 'specified subject.'

- (46) We promised Bill [_s PRO to kill each other].

and:

- (47) *We persuaded Bill [_s PRO to kill each other].

In (46) *each* originates in the deep structure as part of the NP containing *we*, which is the controller of PRO and moves into the subordinate sentence by 'each movement' (cf. 44 above); whereas in (47) the controller of PRO is the NP *Bill* which does not contain *each* at any level of representation. PRO in (47) is therefore a specified subject. Note that any phonologically realized NP (such as *the soldier* in (45)) is a specified subject, and also blocks the application of 'each' movement.

Under standard assumptions, there was an immediate counter-example to these conditions: the rule of

wh-movement (so called because it moves phrases which contain the class of interrogatives and relatives that begin with the letters 'wh,' such as *who*, *which*, and *what* – as well as various others not fitting this description, such as *how*). *Wh*-movement is the rule responsible for deriving sentences such as:

- (48) Who did John kill?

Since *kill* is a transitive verb, it is expected to have a direct object, yet no NP follows the verb in (48). The reason for this is that the requirement for an object NP is satisfied at deep structure, and the NP concerned, *who*, is moved to its surface structure position by a transformation.


In contrast to the transformations discussed so far, *wh*-movement has the property of being 'unbounded.' That is to say, that there seems to be no principled limit on the amount of material that can intervene between the deep structure position of the *wh*-item and its surface structure position, as the following examples illustrate (the underscore marks the position from which *who* has moved):

- (49) Who does Max think John killed____?
(50) Who did Sam say Max thinks John killed____?

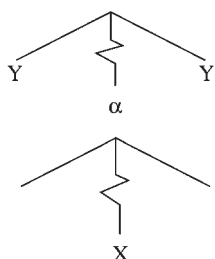
Note that in these examples the clause which contains the extraction site of *who* is a finite one and, in addition, contains a specified subject (*John*). According to the TSC and SSC these sentences should not be possible.

Chomsky's solution to this problem was the radical proposal that, contrary to appearances, *wh*-movement is a cyclic rule. He adopted a proposal made by Bresnan (1970) that *wh*-movement involves moving the *wh*-phrase into the complementizer position of a clause – roughly speaking, the clause-initial position occupied by *that* or *whether*. This position, known as COMP, has the status of an 'escape hatch' for movement; movement to or from the COMP position does not violate the TSC or SSC.

The derivation of (50) would thus proceed as in (51):

- (51) [_s who [_s did Sam say [_s COMP [_s Max thinks [_s COMP [_s John killed ____]]]]]
- 

To enforce the cyclic application of *wh*-movement (and prevent the movement of *who* in (51) from its deep to surface structure positions in one fell swoop), Chomsky proposed a further condition on movement rules: 'subjacency.' A node X is 'subjacent' to another node Y only if Y is higher in the tree than X and there is at most one cyclic node that dominates X and does not dominate Y. This is illustrated diagrammatically below, where α is a cyclic node (that is, NP or S):



The subjacency condition requires that no movement can take place from X to Y unless X is subjacent to Y. Checking (51) will confirm that the combination of COMP as the landing-site for *wh*-movement and subjacency precludes any other derivation than the one indicated.

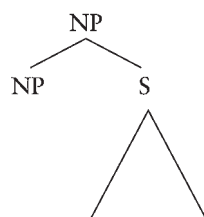
Subjacency also provides an account for a class of phenomena known, since Ross (1967), as 'islands.' Despite the apparent unboundedness of *wh*-movement, it had been known for some time that there are constructions out of which *wh*-movement is prevented for some reason from moving *wh*-items. Chomsky (1964a) had observed that while the sentence *John kept the car in the garage* is ambiguous (meaning either that the car in the garage was kept by John, or that the car was kept in the garage by John), the sentence *Which car did John keep in the garage* is not, having only the latter reading. The nonexistent reading is the one in which the NP *the garage* is contained within the larger NP *the car in the garage*. Chomsky had proposed that in a situation like this, where a transformation might in principle apply to either of two NPs, in fact it must apply only to the larger one. This proposed constraint he termed the 'A-over-A principle.'

Ross (1967) gave a detailed investigation of a wide range of situations in which the application of unbounded movement rules was blocked. Ross pointed out that the A-over-A principle proved too strong in certain cases and identified the defining characteristics of the phenomenon as a set of structural configurations which he termed 'islands,' and proposed a corresponding set of 'island constraints' to restrict the application of unbounded movement rules:

(52) Complex NP Constraint (CNPC)

No item may be moved out of a complex NP by transformation, where a 'complex NP' is a structure of the following form.

Example:



Max met a man that sells books/what

*What did Max meet [NP [NP a man] [_S that sells —]]

(53) Sentential Subject Constraint

No item may be moved out of a sentential subject by a transformation.

Example:

That John gave the book to Max/who surprised Sam

*who did [NP [_S that John gave the book to —]] surprise Sam

(54) Coordinate Structure Constraint (CSC)

No conjunct in a coordinate structure may be moved, nor may any item in a conjunct be moved out of that conjunct.

Example:

Max spoke to Sam/who and Fred

*Who did Max speak to [NP — and Fred]

(55) Left Branch Condition

No NP on the left branch of another NP may be moved out of that NP.

Example:

You read Fred's/whose book

*Whose did you read [NP — book]

The subjacency principle serves to provide a unified account of all these island phenomena except the coordinate structure constraint. Note that in all other cases but coordinate structures, the movement involves a violation of subjacency. To illustrate with the last example, in **Whose* [_S *did you read* [NP — *book*]] two cyclic nodes, S and NP, intervene between *whose* and the extraction site.

Chomsky also proposed that subjacency would account for the prohibition on moving two (or more) *wh*-phrases out of the same clause (the *wh*-island constraint), as in **What* [_S *do you wonder* [*who* [_S — *saw*—]]]. On the assumption that, once the COMP node of the lower clause is occupied by *who*, it is no longer a possible landing-site for *what*, the only route to the topmost COMP will involve crossing two S nodes – a subjacency violation.

Trace Theory

The developments outlined in the preceding sections resulted in a situation where the substantive content of transformations had been severely reduced. Lexicalism had removed many linguistic phenomena from the domain of the transformational component, as had the reanalysis of phenomena

involving coreference; structure preservation reduced the need to specify details in the structural descriptions of transformations, and subadjacency limited the domain of application of transformations.

Chomsky (1973) had introduced the idea that movement of an NP left behind a 'trace' of itself. He developed this idea further (Chomsky 1976). He proposed that all movement transformations leave behind an phonetically empty copy (a 'trace') bound by the moved constituent and inheriting all its relevant properties. He points out the pattern of reciprocal interpretation (now effected by an interpretive rule, rather than the transformation of 'each' movement) supports the postulation of such an empty category in sentences involving raising to subject. There is a parallel between (56) and (57), where reciprocal interpretation in (57) allows *each other* and *the men* to be construed together, just as if they were in the same clause, as in (56):

(56) The men like each other.

(57) The men seem to John [*t* to like each other].

There is a similar parallel between (58) and (59), where reciprocal interpretation is blocked by the SSC. This insight is captured explicitly by the occurrence of a trace of NP preposing in (57) which participates in the interpretation of *each other* and one in (59) which forms a specified subject since it is bound by the singular moved NP *John*, blocking reciprocal interpretation with *men* just as the overt occurrence of *John* does in (58):

(58) The men want [John to like each other].

(59) John seems to the men [*t* to like each other].

Lasnik (1977) drew a parallel between the relationship which exists between traces and their binders and that which exists between PRO and its controller and between a reciprocal and its antecedent: always the binder must precede or asymmetrically 'command' the trace. (A commands B if the first S node that dominates A also dominates B.) This means that movement must always be leftward or to a higher position in a tree. Hence, raising is a possible movement rule, but its inverse, 'lowering,' would not be. Similarly, *wh*-movement must always be to a higher COMP node and never to a lower one.

With these developments, Chomsky proposed a constraint on transformations to the effect that the only categories that may be explicitly mentioned in a structural description are those that are changed (i.e., moved, deleted etc.) by the rule. He also suggests that a further step can be taken in the light of the structure preserving hypothesis, which entails that, for example, an NP can only be moved into another

NP position. 'Thus the rules in question [e.g., passive] reduce to the following formulation: move NP' (Chomsky, 1976: 313).

All other details of the operations carried out by the rule are determined by general principles. The landing site being determined by a combination of the structure preserving hypothesis, subadjacency, and Lasnik's condition.

This move was accompanied by a reinterpretation of the status of conditions like the TSC and the SSC, which ultimately spells the end of transformation grammar as such and its evolution into government-binding theory. At their inception, these conditions were construed as constraints on rules, blocking their application. Chomsky now proposed construing them as conditions on surface structure wellformedness. That is to say, 'move NP' can apply freely, but the resulting representations must satisfy the TSC and the SSC. To take a specific example, *John seems [t to like Bill]* is the result of move NP and the trace is bound by an NP (*John*) as required by Lasnik's proposal. In **John seems [Bill to like t]*, although NP movement has applied, the resulting configuration violates the SSC, on the assumption that the relationship between an NP and its trace is just like the one that holds between a reciprocal and its antecedent. The trace does not stand in the appropriate relationship to *John*, any more than *each other* does to *they* in **They expect [Bill to like each other]*.

In 'On binding' Chomsky (1980) took this development a stage further and proposed that the SSC (renamed the 'opacity condition') and the TSC (renamed first the 'propositional island condition,' or PIC, and then the 'nominative island condition,' or NIC) should be reinterpreted as constraints on the occurrence of anaphors alone, rather than as relations between anaphors and their antecedents. The opacity condition took the form 'if α is an anaphor in the domain of the subject of β , β minimal, then α cannot be free in β , $\beta = \text{NP or } S'$ '. The nominative island condition took the form: A nominative anaphor cannot be free in S' . 'In the domain of' means 'c-commanded by,' where A c-commands B if the first branching node which dominates A also dominates B (a revision of Lasnik's 'precede or command'). 'Free' means not coindexed with a c-commanding antecedent. Nominative case is assigned only to the subjects of finite clauses.

To take up the analysis of *each other* to exemplify the revised framework, Chomsky suggests that all that need now be said about the distribution of *each other* is that it is a reciprocal. From this, within an appropriately articulated theory of grammar, will follow the requirement that it be bound by a plural antecedent within the same S' or NP, if it is not itself

a subject, so that (60) is grammatical with *each other* bound by *the students* within the minimal *S'*:

- (60) John expected [_{S'} the students to like each other].

(61), however, is ill-formed, because *each other* has no acceptable antecedent within *S'*:

- (61) *The students expected [_{S'} John to like each other].

(62) is ill-formed because the lower clause is finite. Its subject *each other* is therefore nominative and, according to the NIC, must be bound in *S'*. But, in this structure there is no possible c-commanding binder within *S'*, so the sentence is defined as ungrammatical:

- (62) *The students expected [_{S'} to each other would to like John].

Earlier, Chomsky (1977) had argued that what had been considered a range of different transformations should all be interpreted as expressions of *wh*-movement. These included not only *wh*-questions and relative clauses, but also topicalization (63a) (and its associated analysis in 63b), *Tough*-movement (64a) and (64b), comparatives and cleft sentences, although most of these do not occur with overt *wh*-phrases in them:

- (63a) This book, I asked Bill to get his students to read.
 (63b) [_{S'}^{TOP} This book] [_{S'}^{COMP} what I asked Bill to get his students to read t]].
 (64a) John is easy for us to please.
 (64b) John is easy for us [_{S'}[who for] PRO to please t]].
 (65a) John is taller than Mary is.
 (65b) John is taller than [_{S'}[what] Mary is t]].

Chomsky's argument was founded on the claim that all these constructions exhibit the defining characteristics of *wh*-movement:

- (66) a. it leaves a gap;
 b. there is an apparent violation of subadjacency, PIC and SSC;
 c. it observes the complex NP constraint;
 d. it observes *wh*-island constraints.

For example, the comparative in (67) and the *wh*-question (68) both demonstrate a violation of the complex NP constraint, while the comparative in (69) and the *wh*-question (70) both demonstrate parallel violations of the *wh*-island constraint:

- (67) *Mary is taller than (what) Bill believes [_{NP} the claim that she is –].
 (68) *Who does Bill believe [_{NP} the claim that she met –].

- (69) *Mary is taller than (what) [I wonder [whether she used to be –]].

- (70) *Who [does Bill wonder [whether she met –]].

The set of transformational movement rules was now reduced to two: 'move NP' and 'move *wh*' and in *On binding* Chomsky proposed that it could be reduced still further to 'move α ,' where α is some category of the grammar. Move α will apply freely, but the resulting structures must satisfy the relevant constraints.

Transformational grammar, as it had been envisaged originally, was about to evolve into a theory in which transformations themselves played a relatively minor role. The task of specifying the operations which transformations were entitled to perform was to be replaced by the task of specifying well-formedness conditions on the representations which resulted from the operation of transformations and other kinds of rules. The only constraint specific to transformations was, and remains, subadjacency.

Transformational grammar constituted a major breakthrough in the investigation of the syntax of human languages and spawned an unprecedented amount of research, providing many new insights into language and its structure. It has been a topic of controversy since its inception, and a variety of alternatives now exist which challenge the very bases of transformational theory; but it is fair to say that many of them have developed in response to positions articulated within transformational grammar and it is doubtful that they would have come about without the stimulus and insights provided by transformational grammar. It is certainly true that our understanding of human language would be much impoverished had it not existed.

See also: Generative Semantics; Principles and Parameters Framework of Generative Grammar; 20th-Century Linguistics: Overview of Trends.

Bibliography

- Bresnan J (1970). 'On complementizers: toward a syntactic theory of complementation.' *Foundations of Language* 6, 292–321.
 Carroll J M & Bever T G (1976). Sentence comprehension: A case study in the relation of knowledge and perception. In Carterette E C & Friedman M P (eds.) *The handbook of perception*. New York: Academic Press.
 Chomsky N (1957). *Syntactic structures*. The Hague: Mouton.
 Chomsky N (1964a). *Current issues in linguistic theory*. The Hague: Mouton.

- Chomsky N (1964b). A transformational approach to syntax. In Fodor J A & Katz J J (eds.) *The structure of language*. New Jersey: Prentice-Hall, Englewood Cliffs.
- Chomsky N (1965a). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky N (1965b). Three models for the description of language. In Luce R D, Bush R R & Galanter E (eds.) *Readings in mathematical psychology*. New York: John Wiley.
- Chomsky N (1970). Remarks on nominalization. In Jacobs R & Rosenbaum P (eds.) *Readings in English transformational grammar*. Waltham, MA: Blaisdell.
- Chomsky N (1973). Conditions on transformations. In Kiparsky P & Anderson S (eds.) *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.
- Chomsky N (1975). *The logical structure of linguistic theory*. New York: Plenum.
- Chomsky N (1976). 'Conditions on rules of grammar.' *Linguistic Analysis* 2, 303–351.
- Chomsky N (1977). On *wh*-movement. In Culicover P, Wasow T & Akmajian A (eds.) *Formal syntax*. New York: Academic Press.
- Chomsky N (1980). 'On binding.' *Linguistic Inquiry* 11, 1–46.
- Dougherty R S (1970). 'A grammar of coordinate conjoined constructions.' *Language* 46, 850–898.
- Emonds J (1976). *Transformational approach to English syntax: Root, structure-preserving and local transformations*. New York: Academic Press.
- Fillmore C (1963). 'The position of embedding transformations in a grammar.' *Word* 19, 208–231.
- Harris Z (1952). Discourse analysis. *Language* 28, 18–23.
- Jackendoff R (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Katz J & Postal P (1964). *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.
- Lakoff G (1970). *Irregularity in syntax*. New York: Holt, Rinehart and Winston.
- Lakoff G (1971). On Generative Semantics. In Steinberg D & Jakobovits L (eds.) *Semantics*. Cambridge: Cambridge University Press.
- Lasnik H (1977). On trace theory. *Linguistic Inquiry* 8, 35–61.
- Lees R B (1966). *The grammar of English nominalizations*. The Hague: Mouton.
- Lees R B & Klima E S (1969). Rules for English pronominalization. In Reibel D A & Schane S (eds.) *Modern Studies in English*. Englewood Cliffs, NJ: Prentice-Hall.
- McCawley J D (1968). Lexical insertion in a transformational grammar without deep structure. *Papers from the Fourth Regional Meeting of the Chicago Linguistic Society*. Chicago, IL.
- McCawley J D (1970). 'English as a VSO language.' *Language* 46, 286–299.
- Newmeyer F J (1980). *Linguistic theory in America*. New York: Academic Press.
- Peters S & Ritchie R W (1973). 'On the generative power of transformational grammars.' *Information Sciences* 6, 49–83.
- Rosenbaum P (1967). *The grammar of English predicate complement constructions*. Cambridge, MA: MIT Press.
- Ross J R (1967). *Constraints on variables in syntax*. (Doctoral dissertation, Massachusetts Institute of Technology.)

Truth Conditional Semantics and Meaning

E Lepore, Rutgers University, Piscataway, NJ, USA

© 2006 Elsevier Ltd. All rights reserved.

From the early 20th century, beginning with the revolutions in logic begun by the German mathematician Gotlob Frege and the English philosopher Bertrand Russell, until the present, philosophers have speculated about the notion of meaning. A variety of such notions are in general use: natural meaning, as in 'smoke means fire' or 'those spots mean measles'; conventional meaning, as in 'a red traffic light means stop' or 'a skull and crossbones means danger'; intentional meaning, as in 'John means well,' or 'Frank means business.' Philosophical semanticists are preoccupied with a different notion of meaning, however: Linguistic meaning is involved in the utterances of 'the words "bachelor" and "unmarried man" are synonymous (have the same meaning)'; 'the word "bank" is

ambiguous (has several meanings)'; 'the string of words "colorless green ideas sleep furiously" is anomalous (has no meaning whatsoever)'; 'the sentence "All bachelors are unmarried" is analytic (true in virtue of its meaning alone)'; and more directly, 'what "La neve e' bianca" means is that snow is white.'

One problem surrounding linguistic meaning concerns its dual roles. The meaning of a word reaches out, as it were, into the world, but also retains an inwards relation to other words. The meaning of 'tomato' is related both to the world – tomatoes – and to other words it combines with, as in, 'Tomatoes are a fruit'; 'Are tomatoes a fruit?' 'Get me a tomato!' Whatever else meaning involves, these two diverse roles are essential. For if one knows the meaning of 'tomato,' one grasps how it applies to the world and also how to employ it in indefinitely many sentences – a phenomenon labeled the productivity (or creativity) of language.

Compositionality is invariably invoked to explain productivity: the meaning of a complex linguistic expression is a function of the meanings of its parts. Differences in meaning between 'A tomato is better than an apple' and 'An apple is better than a tomato' are due solely to the manner in which the sentences are composed. In this regard linguistic meaning differs from other sorts of meaning. A red skull and cross-bones doesn't mean *stop danger*; smoky spots don't signify *enflamed measles*; if John both means well and means business it doesn't follow he means good business.

Finite beings never have entertained nor will entertain or use more than finitely many sentences, but just as 'A tomato is better than an apple' and 'An apple is better than an orange' are sentences, so too is 'A tomato is better than an apple and an apple is better than an orange'; and just as 'The apple is rotten' is a sentence, so too is 'The apple that fell on the man is rotten,' 'The apple that fell on the man who sat under a tree is rotten,' and so on with no obvious end. Because our language includes devices of grammatical conjunction and relative clause construction, we are able to form new sentences from old ones. Whatever meanings are, they must fit together in accordance with the compositionality principle. The project becomes to settle on basic word meaning and to let sentence meaning be composed from simpler components.

Philosophers have tended to fix on one of the two main aspects of meaning to the detriment of other. A theory of meaning – a semantic theory – must explain how a word can perform both of its functions. So, for example, John Locke, the 17th century empiricist, concentrated on the inner aspects of meaning. For him meaning is mental; we use language to encode thought. Successful communication involves correctly decoding words heard into their corresponding associated ideas. So construed, the meaning of an expression is the idea associated with it in the mind of anyone who knows and understands the expression.

But focusing on inner aspects of meaning raises some issues. Suppose your idea of grass is associated in your mind with your idea of warm weather; would it follow that 'grass' has the idea of warm weather as part of its meaning, at least for you? This focus also ignores the public nature of meaning. We learn the meaning of words from others, and once conversant, we can determine whether others successfully understand us. If meaning is an associated idea, how does anyone learn it? Then there is the matter of compositionality. Suppose a speaker associates with the complex expression 'brown cow' the idea of fear; he is not, however, fearful of cows or brown things *per se*, only the brown cows. On an ideational semantics, the

meaning of 'brown cow' is not predictable from the meanings of its parts.

In an effort to render meaning public, B. F. Skinner hypothesized that the correct semantics for a natural language is behavioristic: the meaning of an expression is the stimulus that evokes it or the response it evokes, or a combination of both, at the time it is uttered. The meaning of 'Fire!' might be as it were to run. But a fireman, or a pyromaniac, or an audience of a known liar or a play doesn't run when she hears 'Fire!' Does it seem plausible that such individuals mean something different by 'Fire!' than those who run do? Then too there is the persistent worry over compositionality. Suppose a speaker recoils when he hears 'brown cow' but not when he hears 'brown' or 'cow' alone. The terms' meaning then would be undetermined by the meanings of its parts. How then does a speaker learn its meaning and the meaning of indefinitely many other expressions that can host it?

Bertrand Russell, following J. S. Mill, pursued the intuition that linguistic items are signs of something other than themselves. He suggested that the meaning of an expression is what that expression applies to, thus removing meaning from the minds and bodies of its users and placing it squarely in the world. On a referential semantics, all that one learns when one learns the meaning of 'tomato' is that it applies to tomatoes, and to nothing else. The semantics is compositional: if 'red' applies to anything red and 'tomato' applies to any tomato, 'red tomato' applies to red tomatoes.

But what about words that apply to nothing, like the term 'unicorn'? What about the serious problem first pointed out by Frege, that two expressions may have the same referent without having the same meaning? For example, 'Samuel Clemens' and 'Mark Twain' name the same author, but they aren't synonyms. As Frege noted, one could believe that Mark Twain wrote *Tom Sawyer* yet disbelieve Samuel Clemens did, without being irrational. Some authors, e.g., J. S. Mill and S. Kripke, conclude that proper names lack meaning. After all, if one does not know the referent of a name, one does not reach for a dictionary but rather an encyclopedia. But the problem persists even for common nouns – paradigmatic meaningful expressions. The descriptions 'the 41st president' and 'the husband of Hillary Clinton' apply to the same person but are not synonymous. One can understand both without recognizing that they apply to the same individual. But if meaning is what one learns when one learns an expression and if meaning is reference, then we have a problem.

In the seminal semantic theory of Frege, inner and outer aspects of meaning are inextricably bound.

Frege associates with each expression a referent. The referents of these parts contribute systematically to the truth or falsehood of sentences in which they occur. Thus the truth or falsehood of 'The 41st president is a Democrat' is determined by the referents of its individual words and the way they are put together. Its overall significance is fixed by what the parts of the sentence 'stand for' in the world and by relations between those parts. It follows that if a word is replaced in a sentence with another having the same referent, its truth or falsehood will not change. So the outer condition is secured. But now worries about the inner aspects of meaning arise. If you believe that the 41st president is a Democrat, it doesn't follow you will believe that the husband of Hillary Clinton is one. So 'You believe that the 41st president is a Democrat' will be true, while 'You believe that the husband of Hillary Clinton is a Democrat' will be false. If meaning determines truth or falsity, meaning must involve more than reference.

Frege accounts for this problem with senses. The sense of an expression is, intuitively, not what is referred to by an expression but the way in which it is referred to. Each sense determines one reference, but to one reference there may correspond many senses. Central to his view is that senses are abstract objects and not ideas in people's minds as per Locke. The two belief sentences above can disagree in truth value because, although both are about Bill Clinton, they pick him out in distinct ways.

Frege doesn't address how we explain the reference of words. A natural albeit vague answer is in terms of the psychological capacities of users of a language: words mean what they do because of what speakers do with them. An example of this semantics is the logical positivism of the Vienna Circle in the 1920s and 1930s, according to which the meaning of a sentence is given by an account of what it would take to verify it.

The basic idea behind verificationism is that meaning must be a result of links between language and experience: some sentences have meaning because they are definable in terms of other sentences, but ultimately basic sentences, the observation sentences, have their meaning because of their direct connection with experience (in this case being reports of it rather than copies). The connections between the world and language come down to a connection between observation sentences, on the one hand, and experience, on the other. Speakers, then, refer to whatever objects they would identify as the referents, whether by description or by recognition. Although the circumstances under which 'triangular' and 'trilateral' apply are the same, speakers go about *verifying* those applications in different ways. This suggests

that the meaning of an expression is determined by the verifiability of sentences containing it.

The case against verificationism was most ardently pressed by W. V. Quine in the 1950s. He followed the verificationists in linking meaning to experience, but he argued that experience relates not to individual sentences but to whole theories. Since meaning must be empirically available, Quine poses the question: What empirical evidence determines what certain utterances mean? He contends that the only acceptable evidence is behavioral, based on the demands of the publicity of meaning, and therefore shuns Fregean senses. But behavioral evidence cannot, he argues, determine whether a person's words mean one thing or another; alternative incompatible 'translations' of the evidence are always available. And so Quine infers his famous doctrine that translation is indeterminate: no facts determine what words mean, i.e., there are no meanings.

Confronted with the skepticism about meaning produced by the indeterminacy of translation (and the later Wittgenstein's work as well), Donald Davidson in the 1960s and 1970s made a significant effort to resuscitate meaning. Sharing Quine's sympathies with an outer criterion, Davidson attempted to account for meaning not in terms of behavior but on the basis of truth, which by then had come to seem more logically tractable than meaning due to the formal work of the Polish logician Alfred Tarski. In the 1930s Tarski defined truth for formal languages in terms of the relation of satisfaction holding between the parts of sentences and sequences of objects. Truth is thereby determined systematically by the satisfaction of sentential parts; Tarski could show how to formally derive, from axioms and rules, semantic theorems that specify conditions under which any sentence of the language is true.

Frege and Davidson agree about compositionality being non-negotiable, but Davidson spurns the troublesome notion of sense and instead employs a Tarskian theory of truth to 'serve' as a theory of meaning. In outline, his idea is that a semantic theory for a language should entail for any sentence a theorem that 'gives its meaning' while respecting compositionality. But how can we devise a semantics with these consequences? Davidson's insight was to replace 'means that' in a sentence like "La neve e' bianca" means that snow is white' with 'is true if and only if.' Tarski had already shown how to prove such theorems. By exploiting Tarski's theory of truth as a semantic theory, Davidson rendered substantive the rough but venerable idea that to give the meaning of a sentence is to give its truth conditions.

But how can a truth conditional semantics explain away the phenomena for which Frege invoked senses in order to explain in the first place? The sentences

'The 41st president is a Democrat' and 'The husband of Hillary Clinton is a Democrat' share truth conditions; both are true just in case Bill Clinton is a Democrat. But they don't mean the same. In response, Davidson suggested that in order to devise an adequate truth theory for any given speaker who uses these sentences, we must apply the constraints of radical interpretation, particularly, the 'principle of charity' – assume that on the whole speakers are truthful. Interpretation proceeds as follows: collect sentences a speaker 'holds true,' and devise a truth theory with theorems specifying the circumstances under which these sentences were held true as a consequence. According to Davidson, any such theory will prove a sentence like "The 41st president is a Democrat" is true if and only if the 41st president is a Democrat,' but not "The 41st president is a Democrat" is true if and only if the husband of Hillary Clinton is a Democrat', thus solving Frege's problem.

Most semanticists in the 1980s and 1990s doubted whether a truth theory could specify the meaning of any given sentence. Many recommended adding to a truth conditional approach a conceptual role (also cognitive role, computational role, and inferential role) semantics. To understand how and why, consider the following thought experiment: suppose each of two otherwise psychologically indistinguishable speakers says, 'I am 30 years old.' The references of their utterances of 'I' differ. This shows that if the meaning of 'I' is what each speaker who uses the word grasps, then meaning does not determine reference. For if meaning determines reference, then these speakers do not grasp the same meaning and therefore they must assign distinct meanings to 'I.'

Hilary Putnam and David Kaplan independently explain this phenomenon by distinguishing the character (or stereotype) of an expression from its referent in a context of use. The character of 'I' maps any context of use onto its user. This character is grasped by anyone who understands the sentence. The utterance's content is its truth condition. When each speaker says, 'I am 30 years old,' what they know may be the same – they have the same understanding – but what each says has different truth conditions.

So far two expressions can share the same character (or stereotype or whatever corresponds to what we know in virtue of which we understand) and yet these expressions can differ in reference and truth conditions, since, as we saw above, conversely, sentences with the same truth conditions, our Bill Clinton sentences, can differ on the understanding component of meaning. What are we to make of someone who understands both sentences, yet asserts the first while denying the second? We know Frege's solution appeals to senses, and we touched upon Davidson's effort with

radical interpretation. The conceptual role semanticist argues, instead, that these two descriptions have distinct computational roles for the speaker.

In sum, there are two distinct ways of semantically individuating a speaker's words: according to truth conditions **and** according to computational roles. According to the way of truth conditions, the speaker believes that Bill Clinton is a Democrat and also believes he is not a Democrat. But his rationality is not impugned, since truth conditions do not exhaust understanding. According to computational role, the corresponding beliefs expressed by utterances of these sentences are distinct, since the descriptions have distinct computational roles. When his beliefs are individuated in terms of computational roles, he does not have contradictory beliefs.

While the idea of a conceptual role for an expression has been around in philosophy for some time, what they are and what form a theory of conceptual role is supposed to take is much less clear than the form of a truth theory. For conceptual role semantics to explain how your word 'tomato' and mine can be synonymous, our words must share conceptual roles, but this is extremely unlikely. As long as there is the slightest difference between us with respect to inferences we are prepared to draw from our word 'tomato,' their conceptual roles differ. But then it is difficult to see how sense could be made of communication. If we assign different meanings to 'tomato' because our conceptual rules are distinct, there is nothing in common to be communicated. If we assign the same meaning (and so, assign the same conceptual roles), there is no need for communication. Compositionality is no easier to understand with conceptual role semantics: the inferential roles of complexes need not be determined by the inferential roles of its components – take 'brown cow' again.

For Davidson, belief and meaning are interdependent. One of the lessons he draws is that nothing can genuinely have beliefs unless it also has a public language. Many philosophers have recoiled, both because they think it is undeniable that certain non-linguistic creatures – such as dogs and apes – have beliefs, and because they hope meaning can be explained in terms of beliefs and other mental states. One such influential semantics is H. P. Grice's, who suggested that the meanings of sentences can be reduced to a speaker's intention to induce a belief in the hearer by means of their mutual recognition of that intention.

Grice's analysis of meaning consists of various parts. It begins with a notion of speaker meaning: A speaker S means something by an utterance if and only if S intends the utterance to produce an effect in an audience A by means of A's recognition of this

intention. So, e.g., I discover a person so constituted that, when I tell him that whenever I groan in a certain way I wanted him to wink; thereafter, whenever he recognizes the groan (and with it my intention), he winks. Grice's underlying idea is that I did something ('made a groan') with intentions. First, I intended to produce the utterance (the groan); second, I intended my audience to recognize the utterance; third, I intended that they should recognize my intention to produce the utterance; and fourth, I intended that they should recognize that I intended them to form some new belief (or do some action, etc.). In these circumstances, according to Grice, my groan means to wink. The place of conventional meaning in Grice's conception of language appears to be that it constitutes a feature of words that speakers might exploit in realizing the intentions referred to in the analysis of speakers' meaning.

Although Grice's program is not as popular as it once was, the general idea of reducing meaning to the psychological states of speakers is now widely accepted (*contra* Quine, Davidson, Wittgenstein, and their followers). In this sense Griceans have returned to the 17th century's emphasis on inner

over outer aspects of meaning. How much, in the end, semantic properties can be attributed to features of the human mind remains a deep problem for continued study.

Bibliography

- David K (1989). 'Demonstratives.' In Almog J, Perry J & Wettstein H (eds.) *Themes from Kaplan*. Oxford.
- Davidson D (1984). *Inquiries into truth and interpretation*. Oxford.
- Frege G (1984). *Collected papers*. Oxford.
- Grice H P (1989). *Studies in the way of words*. Harvard.
- Kripke S (1980). *Naming and necessity*. Harvard.
- Locke J (1690). *Essay concerning human understanding*.
- Ludlow P (1997). *Readings in the philosophy of language*. MIT.
- Mill J S (1843). *A system of logic*.
- Putnam H (1975). 'The Meaning of "Meaning."' In Gunderson K (ed.) *Language, mind and knowledge*. Minneapolis.
- Quine W V (1960). *Word and object*. MIT.
- Russell B (1919). *Introduction to mathematical philosophy*. London: New York.
- Wittgenstein L (1953). *Philosophical investigations*. Oxford.

Truth: Primary Bearers

S Goldberg, University of Kentucky, Lexington, KY, USA

© 2006 Elsevier Ltd. All rights reserved.

Introduction

What are the kinds of thing that are capable of being true or false – of bearing a truth-value? This issue in metaphysics connects with central issues in the philosophy of language. As noted in Haack (1978: 79), it has been regarded as bearing on such topics as the semantic paradoxes, the motivation for many-valued logics, and the viability of Russell's theory of descriptions.

To judge by our loose talk, various kinds of thing are capable of bearing truth-value. The obvious candidates are sentences, statements, propositions, and beliefs. Suppose Sally asserts the English sentence (1):

- (1) George W. Bush was president of the United States in 2003.

We might describe her as having made a true statement, as having (assertorically) uttered a true sentence, as having expressed a true proposition, or

as having given voice to a true belief. But supposing that we regard truth and falsity as properties, might it not be the case that 'in the final analysis' only one candidate has these properties nonderivatively?

Consider for example the hypothesis that the truth or falsity of a belief derives from the truth or falsity of the proposition that is believed. This hypothesis can be supported by appeal to the standard analysis of the propositional attitudes. On this analysis, what is shared by, for example, the belief that Jones is in Manhattan and the desire that Jones is in Manhattan is the propositional content *that Jones is in Manhattan*. Now, if the truth of a belief derives from the truth of the proposition believed, we have a natural explanation of the phenomenon whereby the same state of affairs – in this case, Jones's presence in Manhattan – bears equally on the truth of the belief as on the satisfaction of the desire. The explanation would be that the semantic evaluation of the attitudes – their evaluation along the dimensions of truth, or satisfaction, or what have you – is in the first instance an evaluation of their propositional contents. In this way we see how we might support the hypothesis that beliefs have their truth-values derivatively.

The hypothesis that there is a primary bearer of truth-value is the hypothesis that one type of entity bears its truth-value nonderivatively. Which of the three remaining candidates – sentence, proposition, statement – fits this job description?

The Case for the Sentence as the Primary Bearer of Truth-Value

Consider first the sentence. One reason for regarding sentences as the primary truth bearers has to do with a traditional understanding of the aim and scope of logic. Logic is traditionally taken to be (or include) the study of the entailment relation, where a set of sentences of a language *L* entails a sentence *S* of *L* iff *S* is true on every interpretation on which all of the members of are true. Since this talk of ‘truth on an interpretation’ applies to sentences (well-formed formulas), it can be argued that logic itself gives us a motive to treat (atomic) sentences as having their truth nonderivatively (see **Logical Consequence**).

A second consideration in support of the sentence is that certain theories of truth seem to require, or at least to work best with, sentences as the primary bearers of truth. Consider for example those theories on which truth is a relation of correspondence between the facts and the truth bearers (whatever they are). Such theories typically require that candidate truth bearers have a structure that mirrors that of the facts; and it is arguable that sentences are the most apt candidate here (see e.g., Wittgenstein, 1922). So, too, Tarski’s (1944, 1956) semantic theory of truth defines truth in terms of satisfaction, and goes on to characterize satisfaction in a way that makes use of various quasilinguistic distinctions, including that between names and predicates. This, too, appears to call for sentences as the primary bearers of truth.

It is worth noting, though, that neither of the foregoing arguments from the theory of truth is decisive in establishing the sentence as the primary bearer of truth. For example, on at least some conceptions of propositions, propositions can play the roles required by either of the above theories of truth. (For example, the argument from the correspondence theory of truth could be taken to rule out propositions only insofar as these are taken to be unstructured entities: sets of possible worlds, for example.) In addition, there are those theorists who, though rejecting both the sentence and the proposition as the primary bearer of truth, nevertheless endorse Tarski’s semantic theory of truth: one thinks here of Davidson’s (1967) attempt to square Tarski’s theory with a view on which the statement is the primary bearer of truth.

Context and the Question of Truth Bearers

The sentence proposal is not without its share of difficulties. These can be seen when we turn attention from the formal languages typically studied in logic to natural languages. The main problem concerns context sensitivity. A sentence such as (2) is true or false according to who used it, and when.

(2) I am tired

So if sentences are the primary bearers of truth, we face a dilemma. Suppose that sentences which (like [2]) contain a context-sensitive component bear a truth-value. Then such sentences will have to be treated as having a truth-value that changes over time, according to the context of use. But the idea of a truth bearer that changes its truth-value over time is incompatible with the prevalent view (found in Frege, 1918 and elsewhere) that truth bearers have their truth-value timelessly. Suppose then that sentences like (2) do not bear a truth-value. To retain the sentence-as-primary-truth-bearer view, sentences must be found that do bear a truth-value. One natural suggestion comes in the form of Quine’s (1960) ‘eternal sentences’: the truth bearers are those sentences whose truth-value is unaffected by any change of context. But, given other plausible assumptions, the ‘eternal sentence’ view has the revisionary implications, first, that the vast majority of sentences of everyday discourse fail to have a truth-value, and second (and as a result), that most speakers will have a hard time formulating truth-value-bearing sentences.

Presented with this dilemma, the defender of the sentence might think to advert to the distinction between sentence types and tokens. Then the phenomenon of context sensitivity might be taken to establish that it is not sentence types, but rather sentence tokens, that we evaluate for truth. For instance, Mary’s utterance of (2) at noon, and her utterance of (2) at midnight, will present type-identical but token-distinct sentences – in which case we can treat the two sentence tokens as differing in truth-value, while allowing that each token has its truth-value timelessly. (If Mary was tired at midnight of the night of her midnight utterance of [2], then it is **always true** that she was tired then, and her midnight token of [2] is timelessly true.) But it is unclear whether the appeal to the type/token distinction provides a solid defense of the sententialist view of truth bearers. For one thing, it is natural to think that logic deals with sentence types rather than tokens, in which case the sentence token proposal is in tension with the logic-based motivation for the sentence view. For another, it is unclear whether the appeal to the type/token distinction gives us a reason to prefer the sentence

token, as opposed to the statement or the proposition, as the primary bearer of truth. Indeed, it is plausible to suppose that the core lesson of context sensitivity is that it is not sentences themselves (whether construed as types or tokens), but what is expressed by the use of a sentence on a given occasion, that is evaluated for truth.

This understanding of the lesson of context sensitivity has motivated many to look beyond the sentence. Some endorse the proposition, conceived as what is yielded by a semantic evaluation of utterances, as the primary truth bearer. Others, hesitant to endorse propositions given their abstractness and/or the difficulties involved in formulating their individuation conditions, have opted instead for the statement as the primary truth bearer. Since any utterances of e.g., (2) by distinct speakers, as well as any sufficiently temporally distinct utterances of (2) (even by the same speaker), express different statements, treating the statement as the primary truth bearer provides a natural accommodation of context sensitivity. But the relative superiority of the statement to the proposition depends on the controversial claim that statement individuation will be less problematic than, and in any case independent of, proposition individuation.

Ramifications of the Debate about Truth Bearers

As noted at the outset, the dispute over the primary bearer of truth-value bears on various issues in philosophical logic, the philosophy of logic, and the philosophy of language.

Consider for example the class of paradoxes which are often characterized as the 'semantic' paradoxes. The paradigmatic example is the so-called liar paradox, involving a sentence such as the following:

(L) This sentence is false

(For ease of exposition I will speak of sentences, rather than of types or tokens; but the point can be made either way.) If sentences are the primary bearers of truth, we might expect that (L) must be ascribed a truth-value. (This depends on the additional assumption that if sentences are the primary bearers of truth then all sentences must bear a truth-value.) But of course (L) is paradoxical: if it is false then it is true, and if it is true then it is false. Proponents of the sentence-as-primary-vehicle-of-truth view might well take the lesson of the liar to be that we need a logic that permits of more than two truth-values: the third value could be neither true nor false. Alternatively, the lesson of the liar might be that we should reject the claim that it is sentences, as opposed to statements

or propositions, that are the primary bearers of truth: the claim would then be that an utterance of (L) does not express a proposition or make a statement. (Admittedly, both reactions ought to be assessed in terms of the acceptability of their implications, as well as their adequacy as an account of the semantic paradoxes more generally.)

The issue of truth bearers is also relevant to the viability of Russell's theory of descriptions. On the assumption that semantic evaluation is the evaluation of sentences, Russell concluded that sentences containing a definite description – for example, sentences of the form 'The *F* is *G*' – ought to be analyzed as implicitly quantificational, in the manner of 'There is something, *x*, that is *F* and *G*, and anything that is *F* is identical to *x*.' One of the advantages of this analysis is that it can preserve the meaningfulness and truth evaluability of such sentences even if there is no single thing that is *F*. But Strawson (1950) regarded Russell's proposed analysis as plainly artificial: among other things, it requires us to suppose that the grammatical subject of the sentence ('The *F*') is not its logical subject (which, on Russell's analysis, would be the predicates or propositional functions '*x* is *F*' and '*x* is *G*'). On Strawson's diagnosis, Russell was led to such a plainly artificial analysis by his assumption that semantic evaluation is the evaluation of sentences. Strawson rejected this assumption in favor of the view that semantic evaluation is the evaluation of statements; and having done so, he urged that we do not conflate the issue of truth evaluability (a feature of statements) with the issue of meaningfulness (a feature of words and sentences). Thus, on Strawson's analysis, utterances of sentences of the form 'The *F* is *G*' fail to express a statement (and so fail to be truth evaluable) if there is nothing in the relevant context that is uniquely *F*; but even so the sentence uttered would not thereby cease to be meaningful, since meaningfulness is a matter of the rules for the proper use of expressions. In this way we see that issues pertaining to the primary bearer of truth-value have implications for the viability of Russell's theory of descriptions.

The Scope of the Issue

The debate regarding the primary bearers of truth is itself typically presented in connection with the doctrine that truth is a relation (or a property). As it has been formulated here, the issue concerns which entity bears truth-value nonderivatively, and this can seem to be the same as the question regarding which entity nonderivatively possesses the property of truth. One might wonder, then, whether the debate might be dissolved by following so-called minimalist or

deflationary views of truth, according to which truth is not a relation or a property. But as noted in Haack (1978: 83–84), there are questions in the vicinity of the ‘primary bearer’ debate which survive the repudiation of truth as a property or relation. The key issue concerns the application of formal logic to natural language. Consider the familiar *modus ponens* inference rule:

$$\frac{p \quad p \rightarrow q}{q}$$

To apply *modus ponens*, we must find acceptable values for *p* and *q*. But when considering what counts as an acceptable value, we are forced to confront issues of sentential context sensitivity again. To wit: the inference from “I am tired now” and “If I am tired now, I should go to bed now” to “I should go to bed now,” though valid if we keep fixed the time and speaker, equivocates (and so is invalid) if the speaker or time changes from one premise to the next. This has nothing to do with whether one regards truth as a relation; but it does appear to raise most, if not all, of the issues brought up by the primary-bearer debate.

See also: Logical Consequence; Paradoxes, Semantic; Truth: Theories of in Philosophy.

Bibliography

Bar-Hillel Y (1957). ‘New light on the liar.’ *Analysis* 18, 1–6.

- Davidson D (1967). ‘Truth and meaning.’ *Synthese* 17, 304–323.
- Dummett M (1999). ‘Of what kind of thing is truth a property?’ In Blackburn S & Simmons K (eds.) *Truth*. Oxford: Oxford University Press. 264–281.
- Frege G (1918). ‘The thought: a logical inquiry.’ Reprinted in Strawson P (ed.) (1967). *Philosophical Logic*. Oxford: Oxford University Press.
- Haack S (1978). *Philosophy of logics*. Cambridge: Cambridge University Press.
- Kneale W C (1971). ‘Russell’s paradox and some others.’ *British Journal for the Philosophy of Science* 22, 321–338.
- Kneale W C & Kneale M (1962). *The development of logic*. Oxford: Oxford University Press.
- Kripke S (1975). ‘Outlines of a theory of truth.’ *Journal of Philosophy* 72, 690–716.
- Lewy C (1946). ‘How are the calculuses of logic and mathematics applicable to reality?’ *Proceedings of the Aristotelian Society* 20 (supplement), 30–39.
- Quine W V O (1960). *Word and object*. Cambridge, MA: MIT Press.
- Strawson P (1950). ‘On referring.’ *Mind* 59, 320–344.
- Tarski A (1944). ‘The semantic conception of truth.’ *Philosophy and Phenomenological Research* 4, 341–376.
- Tarski A ([1931] 1956). ‘The concept of truth in formalized languages.’ In Tarski A (ed.) *Logic, semantics, and metamathematics*. Oxford: Oxford University Press.
- Wittgenstein L (1922). *Tractatus logico-philosophicus*. London: Routledge and Kegan Paul.

Truth: Theories of in Philosophy

M P Lynch, University of Connecticut, Storrs, CT, USA

© 2006 Elsevier Ltd. All rights reserved.

Traditional Theories

‘What is truth?’ is a question often deemed to be so perplexing as to be unanswerable. Yet understanding truth is clearly important, as the concept seems to sit at the center of many debates, including those over the nature of knowledge, meaning, and logic. Theories of truth are divided over two basic questions: Does truth have a nature? And, if so, what sort of nature does it have? Traditional theories answer ‘yes’ to the first question and so attempt to give an answer to the second; deflationary theories deny that truth has a nature that needs a deep

explanation. We’ll canvass traditional and deflationary theories in turn.

The oldest theory of truth is arguably the correspondence theory of truth. The basic idea, which may have originated from Aristotle, is that true propositions correspond to the facts. Taken just by itself, this idea is a truism; correspondence theorists add flesh to its bones by spelling out the nature of ‘correspondence’ and ‘facts’ in different ways.

In the early 20th century, for example, Wittgenstein (1922) and Russell (1906) developed a version of the correspondence theory of truth according to which propositions were true in virtue of sharing a common structure with the facts. According to this view, the concepts that compose simple propositions like the *cat is on the mat* exhibit a certain form; and objects

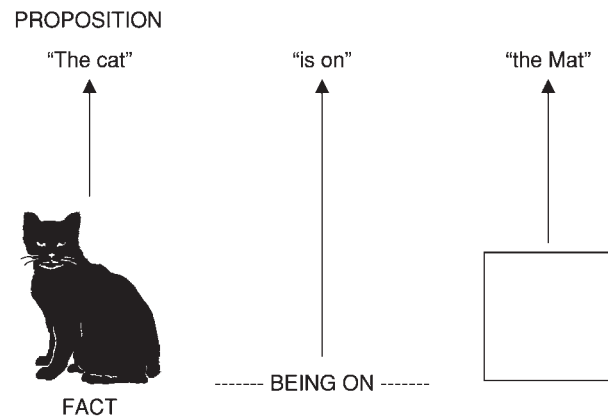


Figure 1 Proposition corresponding to a fact.

(cats, mats) and relations (*being on*) compose basic facts that also exhibit a logical form or configuration. Thus a proposition ‘corresponds’ to a fact just when they share the same form or structure (see **Figure 1**).

Although this sort of view seems to explain the truth of simple propositions about cats and mats well enough, it faces problems with other sorts of facts. Consider the true proposition that there are no unicorns. Is there a negative fact which makes this proposition true? It is unclear whether there are any such facts. In response to these sorts of worries, some philosophers, following Tarski (1943) have attempted to cash out the correspondence intuition in a way that avoids commitment to facts, negative or otherwise, by arguing that propositions are true in virtue of the reference of their parts to objects and properties. The success of this strategy, however, rests on whether a satisfactory account of reference can be given.

Many philosophers also have found fault with the correspondence theory because of worries that it leads to skepticism. According to these philosophers, it is difficult to understand how we could ever “step outside” all language and thought in order to check and see whether our propositions really correspond to the facts. Partly because of these misgivings, philosophers in the 19th century developed the coherence theory of truth (Bradley, 1893). According to the coherence theory, a proposition is true just when it is a member of a coherent system. Thus propositions aren’t true on this view because of their relation to the mind-independent facts but because of their relation to each other. One benefit of this view is that it seems to provide an answer to skepticism. For as long as a proposition ‘hangs together’ with other propositions in the sense defined by the theory, it is true. We don’t have to encounter the naked facts ‘in themselves’ to know what is true and what is false.

Coherence theories, however, have seemed overly permissive to many philosophers. As Russell (1906) famously pointed out, such views imply that any group of propositions – even ones that are obviously false – are true as long as they are mutually coherent. They therefore seem to make too many propositions true.

A third theory of truth was developed by the pragmatist William James (1909) following earlier work by Charles Peirce. According to James’s pragmatist theory, a proposition isn’t true because it copies or coheres with anything. For the pragmatists, propositions are tools, and good tools do good work. Thus, on this theory, a proposition is true just when it is practically justified. A proposition is practically justified in turn when it is useful for achieving our goals over the long run.

The pragmatist theory, however, seems to face an uncomfortable dilemma when asked to explain what makes a proposition useful. If what makes propositions useful is up to us, then again any old claim might turn out to be true. By contrast, if it isn’t up to us, then it seems we must abandon pragmatism. Consider the proposition that my brakes are working in my car. Believing this guides me in making future useful predictions (if I press the brakes the car will stop). But presumably it is only useful for me to believe that the brakes are working if it is really true that they *are* working. But this is not something the pragmatist can say, since we just invoked truth to explain usefulness instead of the other way around.

Deflationary Theories

Deflationary theories of truth are so called because they deflate the pretensions of traditional theories which attempt to define the underlying nature of truth. Deflationists take the fundamental fact about truth to be

(T) The proposition that *p* is true if and only if *p*.

The basic idea is that this schematic principle (more or less, depending on the variety of deflationism involved) explains all there is to explain about truth.

Most deflationary views have two aspects. The first and arguably most central is the metaphysical thesis already mentioned that truth has no underlying nature. This is often put by saying that there is no property, or at least no substantive property, shared by all and only true propositions. Deflationists typically argue for this claim by trying to show that, contra the tradition, we needn’t appeal to anything other than (T) to explain truth’s involvement in

knowledge, meaning and logic. The second aspect of any deflationary theory is a semantic thesis about the role, if any, the word “true” plays in our language.

The simplest form of deflationism is the redundancy theory of truth, and stems from work by Frege and possibly Ramsey (1927). On this account, we need no separate semantic explanation of truth, for ‘it is true that p’ is just another way to say ‘p.’ To say that it is true that roses are red, for example, is really just to say in different words that roses are red. Thus, we only *appear* to ascribe a property to a sentence or proposition when we say that it is true; we are in reality ascribing nothing as ascriptions of truth are simply superfluous or redundant; they add nothing to the content of what has already been said.

The redundancy theory faces two crippling problems. First, there are counterexamples to its central claim. Consider, to take just one example, the proposition that

(S) Everything that Socrates said is true.

Clearly, the claim that *Everything Socrates said is true* is not equivalent to the fragment that *Everything Socrates said*. Accordingly, all ascriptions of truth are not redundant as the theory demands.

The second problem for the redundancy theory and other early forms of deflationism concerns whether and how we generalize over schema (T) in order to offer a suitable definition of truth. Redundancy theorists had a problem offering such a definition because using ordinary objectual quantification, it is hard to see how one could convert (T) into a suitable explicit definition. The natural suggestion ‘For any proposition p, the proposition that p is true if and only if p’ isn’t, strictly speaking, well formed, because ‘p’ is being used as two different types of variables. Where it appears as a conjunct by itself, on the right-hand side of (T), it is used as a proposition; but, on the left-hand side, it has the character of a name being completed by a predicate.

Contemporary deflationists have made numerous suggestions as to how to deal with these two problems. Like the redundancy theory, the minimal theory (Horwich, 1998) holds that (T) is the most fundamental fact about truth. But, unlike the redundancy theorist, minimalists believe that ‘the proposition that snow is white is true’ and ‘snow is white’ are only necessarily equivalent, not synonymous. Moreover, they can acknowledge that truth is property. It is just not a substantive or natural property that needs any sort of *metaphysical* explanation. Minimalists also reject the need for an explicit definition of truth. They argue that truth is instead

implicitly defined by all the instances of (T), and that our grasp of the concept consists in our disposition to accept these instances. Finally, following Quine, minimalists hold that our concept of truth does serve an important purpose. Without it, rather than asserting (S) above, we would face the impossible task of expressing an indefinitely long conjunction, namely:

If Socrates said roses are red, then roses are red,
if Socrates said violets are blue then violets are blue ...

But with (T) in hand, we may say instead that for every object x, such that x = what Socrates said, then x is true. The minimalist concludes that the sole function of the concept of truth is to allow us to generalize over open-ended strings of claims like this.

One issue dividing contemporary deflationists concerns the proper bearers of truth. Disquotationalists (e.g., Quine, 1990) are skeptical about the existence of propositions and take sentences or even utterances, to bear truth. Hence, they take as their central principle not (T) but the disquotation schema

(D): “S” is true if and only if s.

So-called pure disquotationalists, following Field (2001), go even farther. They take the concept of truth to apply only to sentences that we understand. Other philosophers have objected that this is too stringent, as it seems that we can apply the concept to sentences we can’t understand, as when I say that most of what Socrates said was true, even though I don’t understand ancient Greek.

A pressing problem for any form of deflationism, first raised by Dummett (1978), is how to account for the normative role truth plays in thought and language. It is a truism that in asserting propositions and forming beliefs, we aim at the truth. That is:

(TN) Other things being equal, we ought to believe what is true.

But it is difficult to see how (TN), which deals with what ought to be the case, can be deduced from (T), which deals only with what is the case. Deflationists typically reply that (TN) is simply another example of the generalizing function of truth. That is, (TN) is simply short for a long conjunction of individual norms of the form:

(B) Other things being equal, we ought to believe p when p.

And (TN) can be deduced from (T) along with (B). The issue remains unclear, however. For the deflationist now faces the difficult task of explaining why we

should accept all instances of (B) without invoking the fact she set out to explain: that it is good to believe what is true, that is, (T).

Alternative Theories

Recently, several different theories of truth have sprung up as alternatives to both traditional and deflationary views of truth. The identity theory (Dodd, 2000) holds that true propositions do not correspond to the facts but are rather strictly identical to facts. Primitivism (Davidson, 1990) holds that truth is a primitive, indefinable property that we cannot explain in simpler terms that don't already presuppose it. Finally, alethic pluralism (Wright, 1992; Lynch, 2004) holds that truth may come in more than one form, in that propositions in different domains might be true in different ways. Thus, propositions about the physical world around us may be true by corresponding to the facts, whereas those about morality may be true in virtue of their coherence with other propositions. Thus, the pluralist rejects traditional theories in so far as they try to say that all propositions are true in a single way. But, unlike the deflationist, they hold that there are more facts about truth than (T). The main problem with such theories, however, is to spell out exactly what is in common between all these ways of being true; that is, to say what makes them all 'ways of being true.'

See also: Deflationism; Metaphysics, Substitution *Salva Veritate* and the Slingshot Argument; Propositions; Quantifiers; Semantics; Realism and Antirealism; Representation in Language and Mind; Truth: Primary Bearers.

Bibliography

Alston W P (1996). *A realist conception of truth*. Ithaca, NY: Cornell University Press.
Aristotle (1993). *Metaphysics*. Kirwan C (trans.). Oxford: Oxford University Press.

Austin J L (1950). 'Truth.' *Proceedings of the Aristotelian Society, Supp*, vol. 24, 111–128.
Beall J C & Armour-Garb B. *Deflationary truth*. Open Court.
Bradley F H (1893). *Appearance and reality*. Oxford: Oxford University Press.
Davidson D (1990). 'The structure and content of truth.' *Journal of Philosophy* 87, 279–328.
Dodd J (2000). *An identity theory of truth*. London: Macmillan.
Dummett M (1978). *Truth and other enigmas*. London: Duckworth.
Field H (2001). *Truth in the absence of fact*. Oxford: Oxford University Press.
Hill C (2002). *Thought and world*. Cambridge: Cambridge University Press.
Horwich P (1998). *Truth*. Oxford: Oxford University Press.
James W (1909). *The meaning of truth*. New York: Longmans Green.
Lynch M P (2001). *The nature of truth: classic to contemporary perspectives*. Cambridge, Mass: MIT Press.
Lynch M P (2004). 'Truth and multiple realizability.' *The Australasian Journal of Philosophy* 83, 384–408.
Peirce C S (1932). *Collected Papers* (vols 2–4). Cambridge, MA: Harvard University Press.
Putnam H (1982). *Reason, truth and history*. Cambridge: Cambridge University Press.
Quine W V (1990). *The pursuit of truth*. Cambridge, MA: Harvard University Press.
Ramsey F (1927). 'Facts and Propositions.' *Proceedings of the Aristotelian Society, suppl*, vol. 7, 153–170.
Strawson P (1950). 'Truth.' *Proceedings of the Aristotelian Society, suppl*, vol. 24, 125–156.
Russell B (1906). 'On the nature of truth.' *Proceedings of the Aristotelian Society* 7, 28–49.
Tarski A (1943). 'The semantic conception of truth.' *Philosophy and Phenomenological Research* 4, 341–375.
Wittgenstein L (1922). *Tractatus logico-philosophicus*. Ogden C K (trans.). London: Routledge and Kegan Paul.
Wright C (1992). *Truth and objectivity*. Cambridge, MA: Harvard University Press.

20th-Century Linguistics: Overview of Trends

G Graffi, University of Verona, Verona, Italy

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Two theoretical trends can be considered as hallmarks of 20th-century linguistics: structural linguistics and generative grammar. They almost equally

divide this epoch: structural linguistics (or, shortly, structuralism) flourished between the 1910s and the 1950s, generative grammar from the 1950s. Structuralism was not a unitary theory, but rather a galaxy of schools sharing some principles; furthermore, some important differences distinguish European structuralist schools from the American one. Generative grammar, instead, originated as a unitary theory,

which subsequently divided into different schools and which stimulated several alternatives from scholars not accepting it.

Both structural linguistics and generative grammar also had an impact outside linguistics: between the 1950s and 1970s (especially in France), the former became the model for all humanities, hence a 'structural' anthropology, a 'structural' sociology, etc., were developed. Generative grammar, in its turn, was seen as one of the initial steps of the so-called cognitive revolution. Neither of these extensions was free of problems: in many cases, concepts of structural linguistics were applied to other fields with some illegitimate modifications, and the debate on what 'cognitive' really means has not yet come to a solution. All this, however, does not lessen the outstanding role of both structuralism and generative grammar within 20th-century linguistics and within 20th-century thought in general. This article will therefore focus almost exclusively on these two theoretical trends: even sociolinguistics will be dealt with rather as an alternative to generative grammar rather than in its applied aspects (language policy, etc.). Also other fields, such as language teaching, experimental phonetics, and so on, will not be presented in this overview: this does not mean that they have not reached important results during the 20th century.

20th-Century Linguistics vs. 19th-Century Linguistics: Continuities and Breakthroughs

It is a widely held opinion that the 19th century has been 'the century of comparative and historical linguistics' and the 20th century that of 'general' or 'theoretical' linguistics. Such an opinion is certainly not ungrounded, but it needs some qualifications. Indeed, historical linguistics in the modern sense originated in the 19th century and experienced an astonishingly fast development: in the course of about 80 years, the whole structure of historical-comparative grammar of Indo-European languages reached its final form. Later discoveries (e.g., of languages like Hittite or Mycenaean Greek) added some new data, but the overall architecture built by the Neogrammarians nevertheless remained valid, and it is still today the frame of reference for any historical linguist working in the domain of comparative grammar of Indo-European languages. However, historical-comparative grammar was not the only subject investigated by 19th-century linguists; as a matter of fact, many of them dealt with topics that one would certainly label, today, 'general linguistics.' This phrase may refer to somewhat different research perspectives, as, e.g., (1) speculation on language

in general, hence also on language change (and investigation of the principles of historical linguistics plainly enters into this kind of research), and (2) all kind of linguistics that is not historical ('synchronic,' or 'panchronic' in Saussure's terms). Both kinds of general linguistics were practiced during the 19th century. W. von Humboldt (1767–1835) was not an isolated exception: his speculations on the nature of language and his typological classification of languages were developed by several of his followers, such as Heymann Steinthal (1823–1899), Georg von der Gabelentz (1840–1893), and Franz Misteli (1841–1903). But also the first generations of comparative linguists had the study of language in general as their first goal. For example, Franz Bopp (1791–1867) reconstructed Proto-Indo-European verbal forms according to a scheme of verb phrase, which is heavily influenced from Port-Royal views. Even August Schleicher's (1821–1868) views about language and language change belong to 'general linguistics' in the former of the senses alluded to earlier. The debate of the 1880s, about the 'sound laws' (*Lautgesetze*), possibly marks the highest point of this kind of general linguistics: shortly after conclusion of such debate, it gradually became less and less important. A 'paradigm' in the Kuhnian sense has developed: the majority of scholars consider only a given set of problems as 'scientific,' namely those of a historical kind. This paradigm is general labeled as the 'Neogrammarian' one: but it cannot be forgotten that many of the Neogrammarians also dealt with topics of general linguistics in both senses quoted earlier. The often labeled 'Neogrammarian Bible,' namely Hermann Paul's (1846–1921) *Prinzipien der Sprachgeschichte* (I ed. 1880; V and last ed. 1920), deals with topics both of historical linguistics and of general linguistics. For example, Paul defined some oppositions that seem to foreshadow some Saussurean dichotomies (see next section) – that of 'descriptive grammar' vs. 'historical grammar' (which could be held to correspond to that between synchronic and diachronic linguistics), or that of 'individual linguistic activity' vs. 'linguistic usage,' which could be considered analogous to that of *parole* vs. *langue*. Similar distinctions were also introduced by Gabelentz in his *Sprachwissenschaft* (I ed. 1891; II ed. 1901), where three meanings of the term *language* (*Sprache*) are distinguished: (a) 'discourse' (*Rede*); (b) 'a totality of expressive means for any thought'; and (c) 'linguistic capacity' (*Sprachvermögen*), i.e., "a faculty innate to all peoples of expressing thought by means of language." Sense (a) could be made to correspond to Saussure's *parole*, sense (b) to *langue*, and sense (c) to *faculté de langage*. Even if it is quite probable that Saussure knew both Paul's and

Gabelentz's work, such correspondences are more seeming than real, as will be seen in the next section. The fact cannot be overlooked, however, that such matters typically belong to general linguistics. As can be seen from their life dates, neither Gabelentz nor Paul were much older than Ferdinand de Saussure (1857–1913): why, then, are they normally presented in histories of 19th-century linguistics, whereas Saussure is considered as the 'father' of 20th-century linguistics? This is mainly due to the fact that Saussure and his followers dealt with more or less the same stuff, but in a different perspective. Summarizing so far, one could say that topics in general linguistics show a continuity between the 19th and 20th centuries, but the way of looking at them shows a definite breakthrough.

Ferdinand de Saussure

As is well known, Saussure's *Cours de linguistique générale* (Saussure, 1922) was not directly written by him, but it was compiled by two former students, Charles Bally (1865–1947) and Albert Sechehaye (1870–1946), on the basis of the notes from class lectures given by Saussure in the academic years 1906–7, 1908–9, and 1910–11 at the University of Geneva. It is perhaps lesser known that neither Bally nor Sechehaye attended any of these lectures: they simply reworked and systematized the notes that others had passed to them. As a result, their reconstruction is often considered not quite faithful to the authentic Saussurean thought, especially after detailed studies of the handwritten notes by Godel (1957) and their edition by Engler (1967–74). Tullio De Mauro's very detailed and insightful commentary on the *Cours* (published since 1972 together with Saussure's original text) stresses many points of Saussure's original thinking that were more or less modified by the editors. Today the exact knowledge of Saussure's ideas cannot be gained without the support of De Mauro's commentary and/or the attentive reading of Engler's edition. Nevertheless, because only the Bally–Sechehaye edition was available until the 1960s, this text actually influenced the immediately subsequent linguists. Hence, reference will be made in what follows almost exclusively to the Bally–Sechehaye edition.

Saussure's linguistic views are standardly epitomized by his so-called dichotomies: (1) *langue* vs. *parole*; (2) synchrony vs. diachrony; (3) signifier vs. signified (*signifiant* ~ *signifié*); and (4) associative vs. syntagmatic relations. (1) opposes the social aspect of language, the code shared by a speaking community (*langue*), to the individual speech act (*parole*); (2) opposes the state of a language at a given moment of its history to

its change during the time; (3) defines the 'two sides of the linguistic sign,' namely the 'acoustic image' and the 'concept'; and (4) opposes the relation between elements in succession, such as *teach* + *ing* in *teaching*, to that between elements alternative to each other: e.g., *teaching* as alternative to *learning*, or *studying*, etc. All these dichotomies were surely attested in 19th-century linguistics: that of the social vs. the individual aspect of language was already sketched among others by Paul and Gabelentz (see previous section; it must be added that Saussure also hinted at the notion of *faculté de langage*, language faculty, which shows analogies with Gabelentz's *Sprachvermögen*); the synchrony/diachrony opposition could have a foreshadowing in Paul's distinction between 'descriptive grammar' and 'historical grammar'; the idea that the linguistic sign is two sided, in the sense that the meaning is not external, but internal to it, could be traced back even to the Stoics; and also the existence of two kind of relations in language could already be found, e.g., in some of Paul's pages. Two features, however, strongly differentiate Saussure from his predecessors: (1) a systemic approach, and (2) a strong tendency to define the basic concepts of linguistics without anchoring them to other disciplines, such as sociology, psychology, etc. Saussure's key notion is that of language (*langue*) as a 'system of signs,' each of which has no intrinsic value, but whose value is determined solely by their relationships with the other members of the system ("*dans la langue, il n'y a que des différences*"; Saussure, 1922: 166; original emphasis). This sign system is the code shared by all members of a linguistic community, and its only root lies in this common sharing, because linguistic signs do not have any intrinsic value. Linguistics is a part of a more general science called semiology, namely "*une science qui étudie la vie des signes au sein de la vie sociale*" (Saussure, 1922: 33; original emphasis). *Langue* is therefore a social notion because of its character of semiological code: *parole* is the use by the individual of this general code. Because linguistic signs have no foundation in external reality, but are purely differential entities, they can change across the time; this is the reason for the distinction between synchrony and diachrony. In Saussure's view, however, system exists only in synchrony, at a given moment of time; diachrony does not concern systems, but only isolated elements. A linguistic change is therefore isolated and fortuitous; only when a given sign is changed, a new system is formed, because the relations between signs are different from earlier. Finally, associative and syntagmatic relations are synchronic, because they are essentially systematic.

Saussure's view of language paved the way to what was later called structural linguistics. Even if neither 'structure' nor 'structural' (but just *système*)

occur throughout Saussure's text in a technical sense, the systemic approach to language and the definition of linguistic notions and categories on a purely linguistic basis (i.e., without reference to psychological categories, and so on) became the starting points of structural linguistics.

Saussurean Trends in Europe

Geneva School

The editors of Saussure's *Cours*, namely Bally and Sechehaye, were only weakly influenced by the systemic and autonomous approach to language developed in that book. This may seem paradoxical, but it has to be kept in mind that both Bally and Sechehaye had completed their linguistic formation before Saussure's lectures in general linguistics. Among Saussurean notions, Bally especially deepened the *langue/parole* opposition. According to him, *langue* preexists *parole* from a *static* point of view. This relationship, however, is reversed from the *genetic* point of view, because *parole* preceded *langue* in the genesis of language (see, e.g., Bally, 1965). Sechehaye was concerned with problems of general linguistics and its relationships with psychology and sociology since his first book (Sechehaye, 1908); in subsequent years he analyzed some fundamental problems of syntax, such as the notion of sentence. These analyses are insightful, but essentially extraneous to the structuralist trend. A stronger structuralist approach characterizes the work of Bally's pupil, Henri Frei (1899–1980); he revisited Saussure's ideas on syntagmatic and paradigmatic relationships, on the one hand, while on the other hand he confronted other descriptions of syntactic structure, mainly those of Bloomfield and of the American structuralist school (see later discussion).

Prague School

The so-called Prague School was formed by a group of linguists belonging to the 'Prague linguistic circle,' founded in 1926 by the Czech anglicist Vilém Mathesius (1882–1945). Members of this circle were, among others, the Russian scholars Serge Karcevskij (1884–1955), Roman Jakobson (1896–1982), and Nikolaj S. Trubeckoj (1890–1938). Prague School dealt with a variety of topics, from English syntax to literary criticism; however, it is especially known for its critical development of Saussurean notions expounded in the theses presented to the first International Congress of Linguists ('Prague Theses,' 1928) and for the contributions of some of its members to phonology (especially Trubeckoj and Jakobson).

Among the most influential statements contained in Prague Theses, the following two can be quoted: (1) language is a functional system, whose goal is communication; and (2) the synchrony/diachrony dichotomy is not so neat as Saussure presented it: on the one hand, no linguistic state can be considered as totally independent from evolution and change; on the other, phonetic change is not blind and unsystematic as Saussure assumed, but it must be considered in the framework of the sound system that underpins it.

The most important Prague work on phonology is surely Trubeckoj's posthumous and unfinished *Principles of phonology* (Trubeckoj, 1939). Phonology was defined by Trubeckoj as "the science of sounds of *langue*," whereas phonetics is "the science of sounds of *parole*." The key notion of phonology is phoneme. Phoneme (the phonological unit) is opposed to sound (the phonetic one): the former is abstract, the latter is concrete and 'realizes' the phoneme. When two sounds occur in exactly the same positions and cannot be changed without a change in the meaning of the words, they are different realizations of the same phoneme ('rule 2' of Trubeckoj, 1939): e.g., English [t], [p], and [k] realize three different phonemes, /t/, /p/, and /k/ because they distinguish, among others, the three words *tin*, *pin*, and *kin*. Two sounds may be different and nevertheless belong to the same phoneme: for example, English [t], [p], and [k] in the preceding examples are produced with an extra puff of air, but this puff of air does not occur, e.g., in *spin*. These two sounds do not distinguish any meaning: they are variants of the same phoneme. Phoneme inventories differ from language to language: sounds that realize the same phoneme in a language may be realizations of different phonemes in other languages. For example, aspirated stops are variants of the same phoneme in English, but they realize a phoneme of its own in Hindi.

Since the late 1940s, Jakobson remarked that phonemes are not the 'smallest distinctive units,' but they are actually constituted by even smaller entities, the distinctive features. For example, /d/ differs from /n/ (cf. *dine* vs. *nine*) because of the feature 'nasality'; and it differs from /t/ (cf. *do* vs. *to*) because of the feature 'tensedness.' During the 1950s, Jakobson, together with Morris Halle (b. 1923) further worked out his theory: any phoneme of any language is analyzed as containing or not containing a given feature from a universally fixed set of 12 (later 14) features, whose values are + or –; (binary value, hence the label of binarism given to the theory). For example, English /t/ would have the following features: [-vocalic], [+consonantal], [-compact], [-grave], [-nasal], [+tense], [-continuous] (for the meaning of these

terms, see Jakobson and Halle, 1956). Both consonants and vowels are defined on the basis of the same features, and all languages have only this inventory of features at their disposal (but some languages exploit only some of them). Jakobson's binarism was adopted (with some modifications) also by generative phonology (see discussion under Generative Phonology Section).

Copenhagen School

The most well-known linguists of the 'Copenhagen school' are Viggo Brøndal (1887–1942) and Louis Hjelmslev (1899–1965). Both scholars vigorously maintained the structuralist point of view and their Saussurean heritage. Nevertheless, their approach to language in general and to syntax in particular shows many differences; whereas Brøndal considered *language* to be based on logic, Hjelmslev's program was to give *linguistics* a logical basis, in the sense of the 'logic of science,' which was being developed in the 1930s by the neopositivistic philosophers.

In his most important book, Hjelmslev (1943) aimed at constructing a deductive theory of language that he dubbed *glossematics*. Such a theory should be based on purely linguistic notions ('immanent linguistics') and should follow rigorous methodological standards (it has to be "self-consistent, exhaustive and as simple as possible"). Hjelmslev assigned special importance to Saussure's statement that "language (*langue*) is form, not substance." He therefore distinguished form and substance both on the phonological and on the syntactico-semantic level (expression plane and content plane respectively, in his terms). On the expression plane, a continuous stretch of sound can be differently articulated, according to the different phonemic inventories: e.g., where English distinguishes three nasal phonemes (/m/, /n/ and /ŋ/), Italian only distinguishes two (/m/ and /n/). On the content plane, the same continuum can be differently subdivided: e.g., English divides the color spectrum from green to brown in four sections (*green, blue, gray, brown*), whereas Welsh divides it in three (*gwynedd, glas, llwyd*). Both planes are analyzable, according to Hjelmslev, into smallest units, which are limited in number, that he called *figuræ*: expression *figuræ* are phonemes, content *figuræ* are semantic units from which larger semantic units can be constructed (e.g., *man* would be formed by the content *figuræ* 'human,' 'male,' 'adult'). Content *figuræ* and expression *figuræ* are not in one-to-one correspondence: this is the reason why two planes are postulated (otherwise, such a postulation would be superfluous and the theory would violate the simplicity requirement). Any structure that has an expression

plane and a content plane is named by Hjelmslev a *semiotic*, whereas structures with one plane only are 'symbolic systems.' Each plane can in its turn be constituted by a *semiotic*, and so on.

Structural Linguistics in France: Benveniste, Martinet

Émile Benveniste (1902–1976) combined his experience in the field of historical-comparative grammar of Indo-European languages with a particular skillfulness in the analysis of linguistic facts. He was surely well acquainted with the patterns of investigation worked out by European and American structuralism, but he resorted to them only to a limited extent. Somewhat paradoxically, this allowed him to sketch some analytical proposals that are sometimes superior to those of the structuralists. Among such proposals (see Benveniste, 1966), the most well known are his remarks on Saussure's notion of arbitrariness of linguistic sign and those concerning the definition and classification of grammatical persons and of pronouns, which parallel the investigations about performative utterances developed by Austin more or less during the same years (cf. Pragmatics section).

André Martinet (1908–1999) was in the 1930s, a foreign member of Prague linguistic circle; he consistently developed that 'functional view' of language explicitly stressed by the Prague theses (see Prague School section). Natural languages, in Martinet's view, have three features in common: (a) their communicative function, (b) their use of vocal utterances (i.e., natural language is essentially and primarily a vocal phenomenon, and only derivatively a written one), and (c) the double articulation, i.e., a first articulation into significant units ('monemes,' a term borrowed from Frei, but with a somewhat different sense, to replace 'morpheme'), which are in their turn articulated into distinctive units ('phonemes'). One of the most interesting instantiations of Martinet's functionalism is his investigations of diachronic phonology (cf. Martinet, 1955); the 'economy' of sound changes is the effect of the balance of two opposed tendencies, the 'minimal effort' (which tends to lessen sound differences) and 'communicative efficiency' (which tends to multiply them).

Other European Scholars (Guillaume, Tesnière, London School)

The scholars presented in this section, however, certainly influenced by Saussurean thought and hence ascribable to the structuralist trend, nevertheless remained somewhat aside from the debate that developed about the basic tenets of structural

linguistics, especially between the two World Wars. This fact occurred also because the most significant works of some of them (e.g., Guillaume and Tesnière) were only posthumously published (see Guillaume, 1971–1990; Tesnière, 1966).

The writings of the French linguist Gustave Guillaume (1883–1960) are often difficult to read and interpret, especially because of the dark philosophical style that shows many influences. That by Henri Bergson (1859–1941) is especially significant in Guillaume's analysis of concept of 'time' (which he does not clearly distinguish from that of 'tense'). Among the several topics dealt with by Guillaume, one can quote his opposition between what is a formal expression in language ('psychosemiotics') and what is expressed by it ('psychosystematics'); only what has a morphophonological representation of its own can be called semiotic. Guillaume also proposed to replace Saussure's terminological pair *langue/parole* with *langue/discours* (speech). Speech necessarily presupposes language: their relationship can be expressed in terms of the pair 'power' (language) vs. 'effect' (speech).

Lucien Tesnière's (1893–1954) work is especially important for its contribution to syntax. The seminal notion of Tesnière's syntax was that of valency. Tesnière compared the verb to "a kind of hooked atom" that can exert its power of attraction on a smaller or bigger number of 'participant roles' (*actants*). Besides participant roles, the sentence may also contain some 'circumstantial roles' (*circonstants*), which express the conditions of place, time, manner, etc., in which the process described by the verb takes place. Participant roles are obligatory; circumstantial roles are optional. The number of participant roles varies according to the verb class to which the verb belongs, so we have several verb classes according to their 'valency sets.' If Guillaume's linguistic thought did not exert any special influence on subsequent scholars, Tesnière's lies at the origin of several of the most important developments in syntax during the second half of 20th century (cf. Functionalist Schools section).

Among the linguists of the 'London school,' Daniel Jones (1881–1967) and John R. Firth (1890–1960) especially have to be cited. Jones was a phonetician deeply involved in practical questions (such as, for example, the assessment of the principles of phonetic transcription), but he also faced theoretical questions, such as the definition of phoneme. In contrast with the more abstract view held by Trubeckoj (cf. Prague School section), who defined phoneme on an exclusively linguistic basis, Jones opted for a 'physical' definition of phoneme as "family of sounds related in character no member of which occurs in the same phonetic context of any other member." Jones

also rejected Trubeckoj's sharp opposition between phonetics and phonology.

Firth's view of language is characterized by the key role it assigned to the notion of context. He defined 'meaning' as 'function in context': not only words and sentences, but even phonetic units have meaning. Firth's contextual approach was especially fruitful in phonology. In his view, phonology cannot be limited to the segmentation and classification of sounds and phonemes ('paradigmatic units'), but must take into account also prosodic, 'syntagmatic' units such as the syllable (hence the name of 'prosodic phonology' given to Firth's theory). It is therefore necessary to study syllabic structure in terms of general sound classes, such as C(onsonant) and V(owel), and of their respective positions. Firth also maintained that grammar (i.e., syntax and morphology) and phonology are interdependent, anticipating in this way positions that will be later held in generative phonology (cf. Generative Phonology section).

Among Firth's students, one may cite R. H. Robins (1921–2000), especially for his historical researches on classical and medieval linguistics, and M. A. K. Halliday (b. 1925), whose 'Systemic Functional Grammar' (see Functionalist Schools section), worked out since the 1960s in successively revised versions, had several important applications in many fields, as artificial intelligence, discourse analysis, or language education.

American Linguistics from 1920s through 1960s

Sapir and His Heritage

American linguistics began to show peculiar features, different from those of European linguistics, from the beginning of 20th century. At that time, mainly because of the influence of the anthropologist Franz Boas (1859–1942), American linguists oriented a lot of their research to the study of Amerindian languages. Because such languages were devoid of written tradition, these scholars were automatically led to adopt a synchronic point of view. Furthermore, given the difficulty of applying the notions of Western grammar to such languages, attempts had to be made at describing them in purely formal ways: this led to the development of 'distributional' and 'classificatory' methods, i.e., based on the observation of pure occurrences of forms.

Edward Sapir (1884–1939) was himself an anthropologist and a linguist at the same time and devoted much of his research to Amerindian languages. His theoretical ideas are expressed in his book *Language* (Sapir, 1921), and in several papers posthumously

collected (Sapir, 1949). According to Sapir, language is an ‘overlaid function’ from a physiological point of view, because it is a psychological and symbolic phenomenon. Hence, a purely physiological view of speech sounds, as was typical of late 19th-century experimental phonetics, is untenable: e.g., English *wh*-sound as in *when*, *where*, etc., is physiologically identical with the sound produced blowing out a candle, but the two sounds are essentially different, because only the linguistic *wh*- is ‘placed’ in a system that is composed “of a definitely limited number of sounds.” In this way, Sapir arrived at a psychological conception of phonemes and variants.

Sapir’s classification of grammatical concepts, on which his new approach to language typology is based, also deserves special mention. He classified them into two main groups: concepts that express ‘material content’ and ‘relational’ concepts. Each of both groups is further subdivided into two groups: so one obtains ‘basic concepts’ (1a), ‘derivational concepts’ (1b), ‘concrete relational concepts’ (2a), and ‘pure relational concepts’ (2b). Concepts (1a) and (1b) are mainly semantically (or ontologically) based; they are ‘objects,’ ‘actions,’ ‘qualities’ (1a) and their derivations (1b). Gender and number belong to the group (2a), whereas grammatical relations (subject, object, attribute, etc.) belong to pure relational concepts (2b). The possible combinations of the four groups of concepts brings about Sapir’s classification of languages into four ‘conceptual types.’ Type (A) languages only contain concepts (1a) and (2b); those of type (B), concepts (1a), (1b) and (2b); those of type (C), concepts (1a), (2a) and (2b); those of type (D), all four kinds of concepts.

Among Sapir’s followers the important contribution of Morris Swadesh (1909–1967) to phonemics (a term that corresponds to the European ‘phonology’) is to be noted. The most well-known heritage of Sapir’s thought is, however, the so-called Sapir–Whorf hypothesis, so called after his name and that of his follower, the nonprofessional linguist Benjamin L. Whorf (1897–1941). According to the theory, our vision of the world is heavily conditioned by our language. This hypothesis is today rejected, especially by those linguists, as generative grammarians, who maintain that cross-linguistic differences are actually more apparent than real. On the other hand, generative grammarians have often reevaluated Sapir’s work against Bloomfield’s, because of its psycholinguistic (or mentalistic) approach strongly avoided by the latter.

Bloomfield

Leonard Bloomfield’s (1887–1949) behavioristic approach (see especially chap. 2 of Bloomfield, 1933)

essentially consists of describing language as a chain of stimuli and responses (S-r-s-R): a speech event takes place when a nonlinguistic stimulus (S; e.g., hunger) produces a linguistic response (r) in the speaker (*give me something to eat!*) and such a response in its turn induces a linguistic stimulus (s) in the hearer, which has as a consequence the hearer’s nonlinguistic response (R; e.g., providing food).

Bloomfield’s major contribution to linguistics certainly does not lie in this crudely mechanistic view of language function, but instead in the working out of some analytical tools, particularly in the domains of morphology and syntax. The most influential of such tools is the so-called Immediate Constituent (IC) Analysis (see Bloomfield, 1933: chap. 13). In the classical example *Poor John ran away*, the immediate constituents are *Poor John* and *ran away*. The analysis goes on by partitioning *poor John* into *poor* and *John*, and *ran away* into *ran* and *away*. Furthermore, *away* is also analyzed into *a-* and *way*: the principle of immediate constituent analysis applies to morphology exactly in the same way as to syntax. IC analysis was subsequently deepened and formalized, not only by ‘post-Bloomfieldian’ linguists, but also within generative grammar.

Post-Bloomfieldian Structuralism

Linguists most directly influenced by Bloomfield’s thought and analytical techniques especially developed the operational and distributional features of his conception of language. One of the most typical examples of this methodological trend was the so-called prohibition of mixing levels: phonological analysis must precede grammatical analysis and must not assume any part of the latter.

Among post-Bloomfieldian scholars, the following can be quoted, according to the different linguistic domains: concerning phonology, William F. Twaddell (1906–1982), Bernard Bloch (1907–1965), and George L. Trager (1906–1992); concerning morphology and syntax, besides Bloch and Trager, Eugene Nida (b. 1914); a very important intervention in the domain of IC-analysis is due to Rulon S. Wells (b. 1919); possibly, the leader of the group can be considered Charles F. Hockett (1916–2000), who dealt with all such different fields and also worked on theoretical problems, especially in polemics with generative grammar. The most original and influential among American structuralists is, however, Zellig S. Harris (1909–1992): in the 1940s, he was engaged in the deepening and the formalization of Bloomfield’s analytical techniques (see especially Harris, 1951); in the early 1950s, he worked out the notion of transformation. In Harris’s framework, a

transformation is seen as an equivalence relation between two different sentence-forms: e.g., *Casals play the cello* and *The cello is played by Casals*, or *he met us* and *his meeting us* are ‘transforms’ of each other. The notion of transformation (with important modifications) was to become a cornerstone of generative grammar, especially in its first phases (see later discussion).

Tagmemics and Stratificational Grammar

Tagmemics is the name given to the linguistic theory worked out by Kenneth L. Pike (1912–2000) and his associates and students. It combines both Bloomfield’s and Sapir’s insights, but it trespasses the boundaries of American structuralism in many respects. The Bloomfieldian side of tagmemics lies in its analytical techniques, which resume, deepen, and modify Bloomfield’s. On the other hand, Pike’s approach to language is decidedly and explicitly Sapirean: language is seen as a cultural phenomenon, strictly tied to other cultural manifestations of human life.

Stratificational grammar was developed by Sydney M. Lamb (b. 1929) beginning in the late 1950s. It combines a post-Bloomfieldian approach with some European perspectives, mainly Hjelmslev’s glossematics and Halliday’s Systemic Functional Grammar. The number of assumed strata varies from two to six, according to the different versions of the theory. Mostly, four strata are assumed: semotactics, lexotactics, morphotactics, and phonotactics. Stratificational grammar aims at giving an account of all kind of linguistic processes, i.e., concerning both competence and performance (see later discussion): it shows, therefore, a ‘cognitive’ approach that sharply differentiates it from classical post-Bloomfieldian theories and makes it closer to generative grammar, although it is very distant from this latter theory both in the assumed principles and on many technical aspects.

The Beginnings of Typological Linguistics

Language typology was an important field in 19th-century linguistics, but was rather overlooked during the first half of 20th century, because Sapir’s work on the topic remained an isolated exception. Things began to radically change in the 1960s, especially stimulated by the work of Joseph H. Greenberg (1915–2001). In Greenberg’s perspective (see Greenberg, 1966), a close link is assumed between typology on the one hand and universals on the other. Language universals are no longer exclusively conceived as features that every language *must* possess: to such universals, named by Greenberg ‘unrestricted’ universals, also implicational universals and statistical correlations have to be added. The most well

known instances of implicational universals concern the linear ordering of elements. Greenberg assumed as the bases of his language classification three possible choices: (1) whether a language has prepositions or postpositions (‘prepositional’ vs. ‘postpositional’ languages). (2) The position of the verb (V) with respect to the subject (S) and to the object (O). Of the six theoretically possible positions, only three normally occur: VSO, SVO, and SOV. (3) The order of the adjective with respect to the noun it modifies: A (= AN) vs. N (= NA). Such choices are systematically correlated with each other in an implicational way: this implication can be exceptionless or only statistically significant. An instance of the first case is the statement that if a language shows VSO order, it is *always* prepositional (Greenberg’s Universal 3). On the other hand, Greenberg’s universal 4 is an example of ‘statistical correlation’: if a language has a normal SOV order, it is postpositional “with overwhelmingly more than chance frequency.” Greenberg’s insights caused a tremendous development of typological studies, as will be seen in the Typological Linguistics section.

The Birth and Rise of Generative Grammar

The Origins of Generative Grammar

Generative Grammar (GG) is the label for the linguistic theory developed by the American scholar Noam Chomsky (b. 1928) and his followers; a GG, in Chomsky’s own word, is “a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences” (Chomsky, 1965: 8). Chomsky was a student of Harris (cf. previous section), but he early adopted a ‘mentalistic’ approach to the problems of language and knowledge, highly polemical against the behavioristic one, typical of Bloomfieldian and post-Bloomfieldian linguistics.

The first systematic version of Chomsky’s theory appeared in print in a booklet called *Syntactic Structures* (Chomsky, 1957), which was partly an abstract of a much more voluminous work written in the years 1955–56 and published only 20 years later, with some modifications. The main features shown in this book with respect to the tradition of American structural linguistics were the following ones: (1) the goal of linguistic description is no more seen in the analysis of a given corpus, but in the accounting for the intuitions of the native speaker of a given language (well-formedness of sentences, synonymy, etc.). (2) A sharp distinction is traced between linguistic theory on the one hand and grammar on the other. (3) IC-analysis typical of American structuralism (see previous discussion) is formalized in a system of rules called

Phrase-structure (PS) grammar. (4) PS-grammar is shown not able to adequately account for all sentences of any natural language. For example, it cannot account for the intuitive relation that any English speaker recognizes between two sentences such as *Mary gave a book to John* and *John was given a book by Mary*, or between the latter and *Who was given a book by Mary?* To account for such kind of relations, it is necessary to postulate a further level of rules, called transformations. This notion was borrowed from Harris, but it is rather differently conceived. Whereas, for Harris, it is a relation between sentences, for Chomsky it is a relation between structures. This means that the input of a transformation is a sentence in Harris' framework, whereas in Chomsky's one it is an abstract structure often rather remote from the actual sentence that it underlies. The importance given to the notion of transformation in the early phase of GG had the effect that Chomsky's theory was initially known as transformational grammar rather than as generative grammar (actually, the use of the latter label was rather unsystematic at that time).

The Standard Theory

In the decade 1955–1965, the model of grammar described in the previous section was modified by Chomsky himself and by some of his early associates, such as Charles J. Fillmore (b. 1929), Jerrold J. Katz (1932–2002), Edward S. Klima (b. 1931), Robert B. Lees (1922–1996), and Paul M. Postal (b. 1936). The result of such changes was the so-called (by Chomsky himself) standard theory, presented in Chomsky (1965). The overall structure of the standard model is the following one: PS-rules and lexical insertion rules generate the deep structure both of simple and of complex sentences. The application of transformational rules to deep structure produces surface structures. PS-rules, lexical rules, and transformations form the syntactic component of grammar; deep structures are interpreted by the semantic component, giving the semantic representation of sentences; and surface structures are interpreted by the phonological component, giving the phonetic representation.

In Chomsky (1965), also the 'mentalist' interpretation of linguistic theory, explicitly defined as 'part of theoretical psychology,' was maintained and argued for in detail. Chomsky opposed competence, defined as "the speaker-hearer's knowledge of his language," to performance, which is defined as "the actual use of language in concrete situations." The linguist has to discover "the underlying system of rules" (i.e., the competence) "from the data of

performance" (Chomsky, 1965: 4). A grammar that correctly describes the competence of a native speaker of a given language is said to be descriptively adequate. A linguistic theory is said to be explanatorily adequate if it "succeeds in selecting a descriptively adequate grammar on the basis of primary linguistic data" (Chomsky, 1965: 25). The task of linguistic theory, then, becomes that of accounting for the properties of the LAD (Language Acquisition Device), i.e., the device that allows the child to construct a grammar from among a set of possible alternatives.

Generative Phonology

Generative phonology was discussed in several essays since the late 1950s and found its systematic presentation in Chomsky and Halle (1968). The starting point of generative phonology is that phonology is 'not-autonomous' from syntax: some phonological processes depend on morphological and syntactic structure. For example, the falling stress contour of *blackboard* is opposed to the rising one of *black board* because the former is a compound, hence belongs to the syntactic category N, whereas the latter is a Noun Phrase. Therefore the rules of assignment of stress contour must refer to syntactic surface structure (cf. Chomsky-Halle, 1968: chap. 2.1.). This is the reason why the phonological component is said to 'interpret' the syntactic component (see previous section). This strict interrelation assumed between the phonological and the syntactic level is quite contrary to the prohibition of mixing levels typical of post-Bloomfieldian structuralism (cf. section on this topic; Pike had already criticized this principle). Generative phonology considered the autonomous approach as a basic flaw of structuralistic phonology, both European and American, labeled 'autonomous phonemics': the notion itself of phoneme as conceived in such frameworks was rejected. Generative phonologists, on the one hand, took advantage from some difficulties in assigning which variants to which phonemes that had already been remarked upon within structuralistic phonology; on the other hand, they maintained that the assumption of an autonomous phonemic level often produces a loss of significant generalizations (the classical case was that of voicing of Russian obstruents, brought forward by Halle). Hence, generative phonology does not assume a phonemic level, but only a phonological representation and a phonetic representation. The former representation is derived from syntactic surface structure by means of readjustment rules; the latter is derived from the phonological representation by means of phonological rules, which apply in a given order. Both phonological and phonetic

representations are strings of word and morpheme boundaries and of feature matrices. In such matrices, columns are segments, and rows indicate the value of features. Features of generative phonology only partly overlap with Jakobson's ones (see section on Prague School): their number is higher (about two dozen vs. 12 or 14), and they are mainly defined on an articulatory rather than on an acoustic basis. Features are 'by definition' binary at the level of phonological representation, whereas they are not necessarily binary at the phonetic one. An essential part of generative phonology is the so-called theory of markedness (developing, but also essentially modifying, insights of Prague phonology): features, segments and rules are not on the same plane, but some of them are more natural in the sense that they are more frequent, are acquired by the child earlier than others, etc. This greater or lesser naturalness is accounted for in terms of unmarkedness vs. markedness of the concerned entities and rules.

Since the 1970s, alternative approaches to the strictly segmental or linear model of Chomsky–Halle (1968) have been developed. For example, feature values and segments were no more seen as necessarily in one-to-one correspondence, but it was assumed that in some cases a single feature can extend over more than one segment, and, vice versa, a single segment can subsequently take two opposite values of the same feature (autosegmental phonology). It was also assumed that the domain of application of phonological rules is not only determined by the syntactic surface structure and readjustment rules, but also that the phonological representation has a hierarchic structure of its own, not necessarily coinciding with the syntactic one (prosodic phonology).

The Impact of Generative Grammar

Generative grammar (or, more exactly, generative syntax) aroused great interest among linguists shortly after the publication of Chomsky (1957). This interest became still greater in the subsequent decade, especially after the appearance of Chomsky (1965) and also reached logicians and philosophers of language. Generative tenets were not accepted by everybody: quite the contrary, many of them were sharply criticized. However, the large majority of linguists felt obliged to take a position on them. The following tenets were especially the focus of discussion: (1) The mentalistic view of linguistics (cf. The Standard Theory), which was later called cognitive. (2) The assumption that linguistic theory has to deal with 'an ideal speaker–hearer,' within a 'homogeneous linguistic community': i.e., the social and communicative aspects of language do not influence its structure.

(3) The notion of Universal Grammar (UG), resuscitated by Chomsky (1965) with explicit reference to the tradition of *grammaire générale* starting with Port-Royal. From the early 1970s, UG essentially came to mean what he had earlier dubbed the 'language acquisition device' (LAD; cf. The Standard Theory): it was assumed to be universal since it would be shared by all human beings. (4) The postulation of two different levels of representation (deep and surface structure).

It is therefore possible to investigate the development of linguistic trends grown from the last 1960s according to the position they took with respect to the previously listed generative tenets. (1) Chomsky's cognitive view of linguistics was actually opposite the main structuralist trends, both in Europe and America, which conceived linguistics as an autonomous field. This new view was rejected, or at least dismissed as irrelevant, by some strictly formal approaches, such as Relational Grammar (see later discussion). It was, however, shared by the majority of trends during the last decades of 20th century, but often in a rather different way from Chomsky's. Indeed, although Chomsky simply assumed that to do linguistics *is* to do 'theoretical psychology,' many scholars maintained that linguistic explanations have to be traced back to more general psychological or cognitive factors, or, at least, they must be supported by independent psychological evidence. (2) Chomsky's low evaluation of social and communicative aspects of language contrasted with many earlier linguistic trends, even of the structuralist kind: e.g., Prague school defined language as a 'means of communication.' The view of language as a social phenomenon had been maintained at least since Meillet and it was strongly reaffirmed by scholars such as Uriel Weinreich (1926–1967) as an explicit rejection of Chomsky's views. Other opposition came from the pragmatic approaches to linguistic analysis that were developing within the philosophical tradition. More or less explicitly, all such trends opposed a 'social-communicative' view of language to the 'cognitive' one. (3) A revival of interest in the problem of linguistic universals had been already shown by researches such as Greenberg's; Chomsky's notion of UG clearly developed such interest in an unprecedented way. However, Chomsky's version of UG was not accepted by anybody: the different approaches to language universals were strictly linked with the different views of linguistics as a cognitive science and of relationships between language on the one side and social and communicative phenomena on the other. (4) Also the question of levels of representation was often linked to the problems of linguistic universals: several scholars equated 'deep structure' with UG, and

'surface structure' with cross-linguistic variation. These interpretations were misled, because both 'deep' and 'surface' structure had a specific technical value within a theoretical framework (see, e.g., Chomsky, 1975: 82). Nevertheless, they exerted a not negligible impact even on trends that were very distant from the Chomskyan one. Many of the debates between the different generative schools concentrated on the question if a distinction between whether a 'deep' and 'surface' structure is really necessary and on the nature of the 'deep' level.

In the following sections, trends stemming from generative grammar are distinguished from trends alternative to it. Such a distinction only refers to historical roots: the former trends were worked out by linguists originally (i.e., at the epoch of the standard theory) belonging to the generative group, the latter by scholars outside it. Nevertheless, several trends of the former group eventually became wholly alternative to the generative model.

Trends Stemming from Generative Grammar

Generative Semantics and Its Heritage

Generative Semantics (GS) was worked out between the 1960s and 1970s by scholars such as George Lakoff (b. 1941), James D. McCawley (1938–1999), Paul M. Postal (b. 1936) and John R. Ross (b. 1938). It was sharply opposed to the Extended Standard Theory (EST) by Chomsky and some of his followers. Both approaches shared a realistic view of linguistics and a multilevel approach to syntax, but their way of implementing such ideas was totally different. In their first works, generative semanticists rejected some basic assumptions of the standard theory: according to them, (a) deep structure was a useless concept, and (b) linguistic description must be semantically based. This semantic basis was sought in the reduction of linguistic categories to logical and/or psychological categories: semantic representation should be made to coincide with natural logic. In later works, it was assumed that semantic representation also includes typical semantic and pragmatic categories, such as focus or presupposition.

From the early 1970s, the leading ideas that had characterized the followers of GS were gradually abandoned, and each generative semanticist followed his own way. Lakoff first tried to work out a 'fuzzy' grammar, according to which grammatical categories are not discrete, but form a continuum from the noun at one end to the verb at the other. McCawley moved toward an empirical and somewhat skeptical approach to syntax: contrary to EST, McCawley kept on rejecting

any theory of language acquisition that did not take into account general cognitive properties. From this point of view, Cognitive Grammar by Ronald W. Langacker (b. 1942) and his associates could be considered as a legitimate heir to Generative Semantics.

From the middle 1970s, two linguists formerly belonging to the GS group, David M. Perlmutter (b. 1938) and Paul M. Postal, developed a theory called Relational Grammar (RG). RG completely abandoned the notion of transformation as an operation on hierarchically and linearly ordered phrase markers. It also explicitly rejected any aim at being 'psychologically real.' RG takes grammatical relations as primitives and represents clause structure as an unordered set of constituents that bear grammatical relations to each other. Grammatical relations may change from one level ('stratum,' in RG terminology) to another. Strata are not connected by means of transformations, but of Relational Networks, which show which different grammatical relations the constituents bear at different levels.

Fillmore's Case Grammar was often associated to GS, but it is essentially independent from it, even if both approaches wholly replaced the standard notion of deep structure. In Fillmore's view, the 'basic structure' of the sentence consists of the verb and an array of case relationships (see Fillmore, 1968). By 'case,' Fillmore does not mean a morphological category, but an 'underlying syntactic–semantic relationship.' The elements of the basic sentence structure are unordered.

'One-Level' Approaches to Syntax

Generative Semantics pushed the distance between 'deep' and 'surface' structure to its extreme, by identifying deep structure with semantic representation. RG preserved a multilevel approach to syntax. From the middle 1970s, other linguistic trends originated that took the opposite path, giving up the distinction between deep and surface structure and assuming a single level of syntactic representation. The first systematic proposals in this direction are due to Michael K. Brame (b. 1944). The most successful of such 'one-level' approaches were, however, LFG (Lexical-Functional Grammar) and GPSG (Generalized Phrase Structure Grammar). LFG was initiated by Joan Bresnan (b. 1945), a former Chomsky graduate student, and GPSG by a British scholar, Gerald Gazdar (b. 1945), who was later joined in his research program by other British and American linguists. On the one hand, GPSG and LFG share several assumptions: e.g., both avoid transformations and resort to other techniques to solve problems that standard theory dealt with in transformational terms. On the other hand, they

originated from and developed with rather different goals and concerns. LFG's original goal was the search for a 'realistic' grammar. GPSG was worked out mainly on the basis of formal concerns and had no special interest in building a 'psychologically real' grammar.

From EST to the 'Minimalist Program'

The syntactic theory worked out by Chomsky and his closest associates in the period from the late sixties until now had as its primary goal that of implementing the notion of Universal Grammar: the development of an adequate model of UG was seen as the proper goal of the cognitive view of language. This theory was called, during the 1970s, Extended Standard Theory (EST); in the 1980s, Principles and Parameters Theory (P&P) or 'Government-Binding Theory' (GB-Theory); from the early 1990s, the Minimalist Program (MP). Three works of Chomsky's could be considered the landmarks of each of these three phases: Chomsky (1973) for EST; Chomsky (1981) for P&P; and Chomsky (1995) for MP.

EST's main concern was the definition of restrictions on the functioning and on the format of syntactic rules. The first, decisive, step in this direction was the system of conditions on transformations of Chomsky (1973). More or less in the same period, Joseph E. Emonds (b. 1940) and Ray S. Jackendoff (b. 1945) formulated some important constraints on the format of transformational rules (Emonds) and phrase structure rules (Jackendoff). The great abstractness of all such conditions was assumed to be the proof that they could not possibly have been taught by adults or inductively discovered by the child. They were assumed to belong to Universal Grammar, namely the 'innate biological system' that is "invariant about humans" (Chomsky, 1975: 29). The innateness hypothesis, of course, contrasts with the actual cross-linguistic diversity. The Principles and Parameters approach was the first real effort made within the Chomskyan program to provide a systematic account of cross-linguistic differences. The universal features of language were dubbed principles, and the dimensions along which languages can vary, parameters. For example, the fact that a sentence in any language must have a subject would be a principle: but in some languages (e.g., Italian, Spanish, etc., as opposed to English, French, etc.) the subject may be 'null,' i.e., not phonetically realized. This option is called the 'null-subject-parameter': it has 'positive value' in Italian or Spanish, 'negative value' in English and French. Although principles are innate, the values of parameters are to be fixed on the basis of experience. 'Principles and Parameters' approach stimulated an amount of research much larger

than anything previously done within any other framework connected with generative grammar. In particular, the notion of parameter stimulated cross-linguistic investigation of several languages.

Chomsky, however, was more interested in the depth than in the breadth of explanation (in a sense, more to explanatory adequacy than to descriptive adequacy) and since the early 1990s developed the 'Minimalist Program.' The leading criterion of MP can be considered that of economy, namely resorting to the least possible number of entities and of levels of representation. Therefore, MP disposed of the levels of 'deep' and 'surface structure' and assumed Phonetic Form (PF) and Logical Form (LF) as the only levels of representation. Nevertheless, MP cannot be equated with 'one-level' approaches discussed in the preceding section. In fact, also this last version of Chomskyan generative syntax essentially assumes a very abstract relation between the phonetic and the semantic side of language: PF and LF are related by the computational system, i.e., a transformational apparatus. One of the main goals of MP is just to show *why* transformations exist: *prima facie*, they would seem to be antieconomical. The answer is that they exist to replace uninterpretable features, which are also antieconomical: by so doing, both imperfections erase each other. Natural language is therefore a perfectly economical system, and from this point of view, Chomsky maintains, it is very rare among other biological systems.

Trends Alternative to Generative Grammar

Functionalism Schools

The common feature of functionalism is the assumption that language structure is conditioned by its function as a means of communication. This approach was already taken by some structuralist scholars, such as Martinet, and especially the founder of Prague school, V. Mathesius, who distinguished between the formal (i.e., the grammatical) and the actual (i.e., the communicative) partition of the sentence. Mathesius's insights were taken over by Prague linguists of the subsequent generation, such as František Daneš (b. 1919) and Jan Firbas (1921–2000), who coined the term Functional Sentence Perspective to mean Mathesius's actual partition.

From the 1960s, the most significant functionalist schools developed as an explicit alternative to the formal paradigm of generative grammar: Functional Generative Description (FGD), mainly worked out by Petr Sgall (b. 1926) and his associates, which represents a further stage of Prague School linguistics;

Simon Dik's (1940–1995) Functional Grammar (FG) and Halliday's Systemic Functional Grammar (SFG). FGD proponents did not reject generative grammar as a whole, but maintained that it was too partial as an approach to language; on the other hand, they considered exclusively pragmatic approaches to be partial as well.

Despite their differences, all these functionalist schools share an important common core, the main points of which are the adoption of (a) Functional Sentence Perspective, and (b) a kind of Tesnière's valency grammar, in its original form or mediated through Fillmore's Case Grammar. Hence, their fundamental problem was to work out a device to explain the relationship between the system of Tesnière's roles (or Fillmore's 'deep cases') and the grammatical and communicative organization of the sentence.

Typological Linguistics

Typological linguistics since the early 1970s, mainly developed as an attempt to explain word order correlations stated by Greenberg (see the section titled 'The Beginnings of Typological Linguistics'), and it gradually replaced purely syntactic explanations with semantically and pragmatically based ones. It is therefore independent from generative grammar both in its origins and in its achievements. However, some insights stemming from generative grammar influenced typological studies, especially in the 1970s. For example, Winfred P. Lehmann (b. 1916) started from a syntactic model analogous to that of Fillmore's Case Grammar. He assumed an unordered 'underlying structure,' to be converted into a linearly ordered one by a rule with phrase-structure format: therefrom VO-languages vs. OV-languages would result (cf. Lehmann, 1973).

The most significant development of Greenberg's proposals about word order universals is due to John A. Hawkins (b. 1947), who showed that cases that appear as exceptions in Greenberg's treatment are actually not exceptions, if Greenberg's universals are reformulated in a 'complex' form. An example of such reformulation would be the following: "if a language is SOV, then, if it has AN order, it has also GN order" (cf. Hawkins, 1983: 64).

Two other key notions developed in the framework of typological linguistics, especially by Edward S. Keenan (b. 1937) and Bernard Comrie (b. 1947), are continuum and prototype: categories are no more defined, as in generative grammar, in terms of possessing or not possessing a given property, but as clusters of properties. If all such properties occur, the

concerned category is 'prototypical'; the deviations from the prototype are distributed along a 'continuum.'

Sociolinguistics

The label 'sociolinguistics' was firstly used in 1952 by the American Haver C. Currie (1908–1993), but it became widespread from about the late 1960s. In the last decades, this label ended in indicating a variety of researches, both of theoretical and applied kind, from 'ethnography of speaking' to 'language policy.' Between the 1960s and the 1970s, however, a sociolinguistic trend presented itself as an alternative approach to generative grammar.

The leader of this trend undoubtedly was William Labov (b. 1927), to whom the notion of variable rule is due. Variable rules have the format of PS-rules of generative grammar (more exactly, of *context-sensitive* PS-rules): their application or nonapplication, however, is not *categorical*, but it is conditioned by some probability factors, both of linguistic and extralinguistic (i.e., social, stylistic, regional) kind. By resorting to the device of variable rules, Labov was able to account for the different realizations of the same grammatical phenomenon across different social groups (a paradigmatic case was that of the contraction vs. deletion of the copula *be* in white vs. black American English speakers; see Labov, 1969).

The status of variable rules became a topic of intensive discussion. Do they belong to competence or to performance? Labov initially assumed that they are part of competence, but he eventually (in the 1970s) rejected the usefulness of such distinction. On the other hand, a radical revision of Chomsky's notion of competence had been proposed in 1968 by Dell Hymes (b. 1927), who replaced it by that of communicative competence, which indicates the speaker's ability to use language according to the different social and contextual situations. It can be remarked that, since the EST period, also Chomsky referred to a 'pragmatic competence' interacting with the grammatical one. Hence, the problem is whether grammatical competence is independent or not from 'pragmatic' or 'communicative' competence: Chomsky's answer was affirmative, whereas that of sociolinguistics and pragmaticists, negative.

Pragmatics

From the 1960s, pragmatics presented itself as an alternative to the Chomskyan view of language as a cognitive capacity fully independent from its use. Indeed, the roots of pragmatics lay earlier than generative grammar: the term had been created in 1938 by the philosopher Charles W. Morris (1901–1979) and the research field had as its initiators, between the

1950s and the 1960s, two British philosophers of language, John L. Austin (1911–1960) and H. Paul Grice (1913–1988) (neither of whom, however, used the word *pragmatics*). Austin maintained that speech is action (cf. Austin, 1962). The primary evidence for this is given by the utterances called performative by Austin, such as *I promise you to come*, by means of which I am not only *saying* something, but also *doing* it. Performative utterances are a kind of illocutionary act: examples of illocutionary acts are question, order, etc. According to Austin, a speech act, besides the illocutionary act, consists of the locutionary act (the uttering of given words and phrases) and the perlocutionary act (the intended effect of the speech act on the hearer). This classification of speech acts was partly revised by John R. Searle (b. 1932).

The original motivation of Grice's 'logic of the conversation' (which dates back to essays from the 1950s, eventually collected in Grice, 1989) was to show that there is no real divergence between the meaning of symbols such as \sim , \forall , \exists , etc., of formal logic and their counterparts *not*, *all*, *some*, etc., in natural language: the apparent differences of meaning are due to certain principles governing conversation, the conversational maxims. If I utter a sentence such as *Some students passed the examination*, I am undoubtedly saying the truth, even if *every* student did in fact pass the examination, but the hearer normally interprets it as meaning that only some students, and not all of them, passed the examination. This is due to the fact that I have violated the 'maxim of quantity': "make your contribution as informative as is required" (cf. Grice, 1989: 26). From my violation, the hearer has drawn the conversational implicature that only some students, and not all, passed the examination.

The analysis of speech acts and of logic of conversation are still today at the center of interest of pragmaticists. The interest in conversation also made pragmatics include a good deal of text linguistics, which originally started as a project to extend formal techniques of generative kind to units larger than sentences. In recent decades, text linguistics seems to have been replaced by the more empirical and informal conversational analysis, initiated by sociologists such as Harvey Sacks (1935–1975) or Harold Garfinkel (b. 1929), but later adopted by pragmatics-oriented linguists.

See also: Functionalist Theories of Language; Generative Grammar; Generative Semantics; Principles and Parameters Framework of Generative Grammar; Saussure: Theory of the Sign; Transformational Grammar: Evolution.

Bibliography

- Austin J L (1962). *How to do things with words*. Oxford: Clarendon Press.
- Bally C (1965 [1932]). *Linguistique générale et linguistique française* (4th edn.). Bern: Francke.
- Benveniste É (1966). *Problèmes de linguistique générale*, 1. Paris: Gallimard.
- Bloomfield L (1933). *Language*. New York: Holt & Co.
- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky N (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Chomsky N (1973). 'Conditions on transformations.' In Anderson S R & Kiparsky P (eds.) *A festschrift for Morris Halle*. New York: Holt, Rinehart & Winston. 232–286.
- Chomsky N (1975). *Reflections on language*. New York: Random House.
- Chomsky N (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky N (1995). *The minimalist program*. Cambridge, MA: The MIT Press.
- Chomsky N & Halle M (1968). *The sound pattern of English*. New York: Harper & Row.
- Engler R (1967–74). Critical edition of F. De Saussure, *Cours de linguistique générale*. Wiesbaden: Harrassowitz.
- Fillmore C J (1968). 'The case for case.' In Bach E & Harms R T (eds.) *Universals in linguistic theory*. New York: Holt, Rinehart & Winston. 1–88.
- Godel R (1957). *Les sources manuscrites du Cours de linguistique générale de F. de Saussure*. Genève: Droz.
- Greenberg J H (1966 [1963]). 'Some universals of grammar with particular reference to the order of meaningful elements.' In Greenberg J H (ed.) *Universals of language*, 2nd edn. Cambridge, MA: The MIT Press. 73–113.
- Grice P (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Guillaume G (1971–1990). *Leçons de linguistique*. Paris: Klincksieck; Québec: Presses de l'Université Laval [vols 1–4]. Québec: Presses de l'Université Laval; Lille: Presses Universitaires [vols 5–10].
- Harris Z S (1951). *Methods in structural linguistics*. Chicago: The University of Chicago Press.
- Hawkins J A (1983). *Word order universals*. New York/San Francisco/London: Academic Press.
- Hjelmslev L (1943). *Omkring sprogteoriens grundlæggelse*. København: Munksgaard (English translation by F J Whitfield: *Prolegomena to a theory of language* (2nd edn.). Madison: University of Wisconsin Press, 1961).
- Jakobson R & Halle M (1956). *Fundamentals of language*. The Hague: Mouton.
- Labov W (1969). 'Contraction, deletion and inherent variability of the English copula.' *Language* 45, 715–762.
- Lehmann W P (1973). 'A structural principle of language and its implications.' *Language* 49, 47–66.
- Martinet A (1955). *Économie des changements phonétiques*. Bern: Francke.
- Sapir E (1921). *Language*. New York: Harcourt, Brace & World.

- Sapir E (1949). *Selected writings of Edward Sapir in language, culture and personality*. Mandelbaum D G (ed.) Berkeley & Los Angeles: University of California Press.
- Saussure Ferdinand de (1922 [1916]). *Cours de linguistique générale* (2nd edn.). Paris: Payot. (English translation by Roy Harris. London: Duckworth, 1983).
- Sechehaye A (1908). *Programme et méthodes de la linguistique théorique*. Paris-Leipzig-Genève: Champion-Harrassowitz-Eggimann.

- Tesnière L (1966 [1959]). *Éléments de syntaxe structurale* (2nd edn.). Paris: Klincksieck.
- Trubeckoj [Trubetzkoy] N S (1939). *Grundzüge der Phonologie*. Prague: Travaux du Cercle Linguistique de Prague, 7. (English translation by C. A. M. Baltaxe. University of California Press, 1969).

Two-Dimensional Semantics

C Spencer, Howard University, Washington, DC, USA

© 2006 Elsevier Ltd. All rights reserved.

When we ask whether a sentence is true or false, we are always asking with respect to a particular world. We are typically concerned about a sentence's truth value in the actual world, but we sometimes consider its truth value in other possible worlds as well. Thus the evaluation of any sentence is world-dependent in the sense that whether it is true (in a world) depends on the facts about that world. Context-sensitive sentences are also world-dependent in another quite different sense – what they mean depends on facts about the context, or world, in which they are used. For instance, *it's cold here* uttered in Pakistan means that it is cold in Pakistan, and when uttered in New Jersey means that it is cold in New Jersey. Two-dimensional semantics uses a formal apparatus from two-dimensional modal logic to characterize these two kinds of world-dependence. The two-dimensional framework has been applied to a variety of problems in semantics (indexicals and demonstratives and their interaction with modal operators), pragmatics (presupposition), and philosophy (accounts of the *a priori/a posteriori* distinction and the psychological/functional roles of thought). All of these applications depend on various assumptions, which are not implicit in two-dimensional modal logic, and many of which are controversial.

Modal logic allows that expressions may have different extensions in different possible worlds. For instance, it allows that the objects that satisfy a predicate in one world may differ from those that satisfy it in another. In one-dimensional modal logic, the rule that determines each expression's extension in every world, called its intension, is represented as a function from possible worlds into extensions. The intension of a predicate *F*, for instance, is a function that takes a possible world onto the set of individuals that satisfy *F* in that world, and the intension of a

singular term *t* is a function taking possible worlds to single individuals. Two-dimensional modal logic allows that a single expression may be associated with different one-dimensional intensions in different contexts, or worlds, of use. So it associates a two-dimensional intension with each expression, which is a function from possible worlds to one-dimensional intensions, or equivalently a function from ordered pairs of possible worlds into extensions (see Segerberg, 1973; Aqvist, 1973; van Fraassen, 1977 for expositions of a two-dimensional modal semantics for formal languages).

Since a two-dimensional intension takes pairs of possible worlds onto extensions, it has the resources to represent the two different kinds of world-dependence mentioned above. One of the worlds supplies the contextual elements needed to interpret context-sensitive expressions, and the other world supplies the context of evaluation. I will call the entity that plays the former role the world of occurrence, and the entity that plays the latter role the world of evaluation, although no terminology is standard.

Two-dimensional intensions can be represented in a matrix such as **Figure 1**, which gives a two-dimensional intension for a single expression, *s*. In the leftmost column of **Figure 1**, w^1 , w^2 , and w^3 represent possible worlds considered as worlds of occurrence. In the top row, these same three worlds are considered as worlds of evaluation. Suppose that *s* is the sentence *I am in San Francisco*. In w^1 , Ann is the speaker of this sentence and Ann is in San Francisco. She is also in San Francisco in w^3 , but

	w^1	w^2	w^3
w^1	T	F	T
w^2	F	F	F
w^3	T	T	T

Figure 1 Two-dimensional matrix.

not in w^2 . In w^2 , Beth is the speaker, but she is in London in all three worlds. In w^3 , Carl is the speaker, and Carl is in San Francisco in all three worlds. The cells of this matrix are filled in with truth values, since this is the appropriate extension for sentences. The row corresponding to w^1 tells us the truth value, in w^1 , w^2 , and w^3 , of the sentence s , considered as occurring in w^1 . In w^1 , Ann utters this sentence, so it is true just in the case where Ann is in San Francisco. Accordingly, this occurrence is true in w^1 and w^3 , but false in w^2 , as the matrix indicates. Similarly, the row corresponding to w^2 tells us the truth value in these three worlds of the sentence s , now considered as occurring in w^2 . Since Beth, not Ann, utters s in w^2 , this occurrence is true just in the case where Beth is in San Francisco. Since she is in London in all three worlds, the matrix contains an F in every cell in this row. The third row of the matrix tells us that the sentence s , considered as occurring in w^3 , is true in all three worlds. Since Carl is the speaker in w^3 , and he is in San Francisco in all three worlds, the row contains a T in every cell.

The matrix in Figure 1 only contains three possible worlds. But it is in principle possible to construct more comprehensive matrices, which contain many or all possible worlds. When the associated expression is a sentence, these more comprehensive matrices determine a number of propositions of theoretical interest. First, each row in such a matrix specifies a set of worlds in which the sentence s expresses a truth when it occurs in the world of that row. When the row in question is that corresponding to the actual world, the proposition it determines is the proposition s expresses. Another proposition of some theoretical interest is the set of worlds w in which the occurrence of s in w expresses a truth in w . If the matrix is constructed so that worlds w – w^n are arranged in numerical order on the top row and down the leftmost column, this proposition will be the diagonal of the matrix.

Applications

Different philosophers have modified this basic application of the two-dimensional framework in different ways for different purposes in semantics, pragmatics, philosophy of mind and language, and epistemology. Some philosophers follow the model of two-dimensional modal logic and offer a two-dimensional semantics for expressions (Kaplan, 1989). Others apply the two-dimensional apparatus to utterances (Stalnaker, 1978), or thoughts (Chalmers, 1996; Jackson, 1998) instead of or in addition to expressions themselves. Philosophers have also interpreted what I have called the world of occurrence and world

of evaluation in different ways. Some construe both as possible worlds (Stalnaker, 1978). For others, they are ordered n-tuples of contextually supplied elements (Kaplan, 1989) or maximal epistemic possibilities (Chalmers, 2004). Some philosophers think that the worlds of evaluation and worlds of occurrence are the same type of entity, and for others, different kinds of entities play the roles of worlds of utterance and worlds of evaluation within the same application of the two-dimensional framework (Kaplan, 1989; Chalmers, 1996, 2004).

David Kaplan has used the two-dimensional framework to provide a semantics for indexical and demonstrative pronouns. For Kaplan, indexicals and demonstratives are associated with characters, which are rules that determine their contents in different contexts of use. For instance, the character of *I* is the rule that says that an occurrence of *I* in a world of occurrence w refers to the speaker of that world of occurrence, and this individual is the content of this occurrence of *I*. Kaplan identifies worlds of occurrence (which he calls contexts of use) and worlds of evaluation (which he calls circumstances of evaluation) as ordered n-tuples of a speaker, time, location, and world.

Robert Stalnaker has suggested that the two-dimensional framework can explain how utterances acquire pragmatic effects. A pragmatic effect of an utterance is something that the utterance conveys beyond its semantic content, and which is determined both by the sentence uttered and the conversational context in which the utterance takes place. The conversational context supplies what Stalnaker calls a context set, which contains all beliefs and assumptions mutually held by all parties to the conversation, and against which the utterance is interpreted. Stalnaker proposes that we construct a two-dimensional matrix for a specific utterance and include in the matrix all and only those worlds in the context set. Typically, an utterance will express the same proposition in every world in the context set, because this set includes information about what the expressions in the utterance mean, how context-sensitive terms are to be interpreted, and which objects they refer to. However, in some cases the context set will not include all of this information, so that the utterance will express different propositions in different worlds. For instance, consider an utterance of *I am Bill Clinton*. This utterance either expresses or a necessarily true proposition, if the speaker is in fact Bill Clinton, or a necessarily false proposition otherwise. In neither case would the literal content of this utterance serve the typical purpose of ruling out some possibilities that the preceding parts of the conversation had left open.

Yet the utterance can serve this typical purpose if some participants in the conversation do not already know who the speaker is. In such cases, the context set for the conversation will include worlds in which the speaker is Bill Clinton, and worlds in which he is someone else. Thus the speaker's utterance of *I am Bill Clinton* expresses different propositions in different worlds in the context set. Which of these is the one the utterance conveys? Stalnaker suggests that it may be the diagonal of the two-dimensional matrix associated with this utterance. This is the contingent proposition that the speaker of a particular utterance is Bill Clinton, and this is intuitively the information conveyed by this utterance.

Stalnaker and Kaplan, among other philosophers, have also suggested that the two-dimensional framework does some strictly philosophical work. First, it appears to offer some account of the special epistemic properties of sentences like (1) (Kaplan, 1989; Stalnaker, 1978; but see Stalnaker, 2004).

(1) *I am here now.*

Uses of (1) express contingent truths, but (1) expresses a truth in every world of use. So (1) is in a certain sense an *a priori* truth – one need not know anything about the world of use to know that (1) expresses a truth in it. Although the different propositions it expresses in different worlds are all contingent, the diagonal of the two-dimensional matrix it determines will be necessary.

Martin Davies and Lloyd Humberstone have also applied the two-dimensional framework to a similar end, to explain a distinction between superficial and deep contingency earlier introduced by Gareth Evans. (Davies and Humberstone, 1980; Evans, 1979). Evans claimed that a sentence *s* is superficially contingent if and only if there exists a world *w* such that it is not the case that *s* is true in *w*, and *s* is deeply contingent if its truth depends on a contingent feature of the actual world. Typically, deep and superficial necessity coincide, but Evans argues that they need not always do so. Evans introduced this distinction to criticize Saul Kripke's claim that some propositions, such as (2) are both contingent and *a priori*.

(2) *The standard meter stick in Paris is one meter long.*

(2) is contingent because that very stick might have been less than one meter, but *a priori* because one can know (2) without appeal to experience. Evans argued that in all putative cases of the contingent *a priori*, the two notions of contingency come apart, and all such cases are merely superficially contingent. Davies and Humberstone apply Evans's basic distinction to

necessity rather than contingency, and claim that a statement *s* is deeply necessary if and only if the diagonal of the two-dimensional matrix for *s* contains every possible world, and it is superficially necessary if and only if the horizontal of the matrix for the row corresponding to the actual world contains every possible world. They do not attempt to use the two-dimensional framework to characterize *a priori* directly, though they do note that they know of no examples of *a priori* truths that are not deeply necessary. Since they also acknowledge some examples of deeply necessary *a posteriori* truth, they do not identify deep necessity, in their sense, with *a priori*.

David Chalmers and Frank Jackson have made the more ambitious identification of necessity on the diagonal with *a priori* (Chalmers, 1996; Jackson, 1998). Jackson in particular has suggested that there is only one notion of necessity and that the deep necessity Davies and Humberstone have defined is really *a priori* in disguise. For Chalmers and Jackson, thoughts as well as sentences have two-dimensional intensions, which they understand as functions from ordered pairs of centered worlds and uncentered worlds to truth values. A centered world is just a possible world with an individual and a time marked as its center. This two-dimensional intension determines both a primary and a secondary intension. The primary intension of a thought or sentence token *t* is a set of centered worlds *w* in which *t*, considered as occurring in *w*, is true in *w*, and would be found along the diagonal of a two-dimensional matrix. Its secondary intension is a set of uncentered worlds *w* in which *t*, considered as occurring in the actual world, is true in *w*, and would be found along the horizontal row associated with the actual world in a two-dimensional matrix. Chalmers and Jackson argue that primary intensions provide the following robust characterization of *a priori* knowledge: a thought is *a priori* if and only if its primary intension is necessary, that is, it contains every centered world (Jackson, 1998; Chalmers, 2004). Critics have attacked their view for what they see to be its revival of descriptivism about reference and its commitment to internalism about content (Soames, 2005; Byrne and Pryor, forthcoming; Stalnaker, 2004). Although Chalmers and Jackson embrace internalism about thought content, they deny that their application of two-dimensional semantics commits them to descriptivism.

See also: A Priori Knowledge: Linguistic Aspects; Analytic/Synthetic, Necessary/Contingent, and a Priori/a Posteriori: Distinction; Character versus Content; Dthat; Modal

Logic; Possible Worlds: Philosophical Theories; Temporal Logic.

Bibliography

- Aqvist L (1973). 'Modal logic with subjunctive conditions and dispositional predicates.' *Journal of Philosophical Logic* 2, 1–76.
- Byrne A & Pryor J (forthcoming). 'Bad intensions.' In Macia J & Garcia-Carpintero M (eds.) *Two-dimensional semantics*. New York: Oxford University Press.
- Chalmers D J (1996). *The conscious mind*. New York: Oxford University Press.
- Chalmers D J (2002). 'On sense and intension.' *Philosophical Perspectives* 16, 135–182.
- Chalmers D J (2004). 'Epistemic two-dimensional semantics.' *Philosophical Studies* 118(1–2), 153–226.
- Chalmers D J (forthcoming). 'Two-dimensional semantics.' In Lepore E and Smith B (eds.) *The Oxford handbook to philosophy of language*. New York: Oxford University Press.
- Davies M & Humberstone L (1980). 'Two notions of necessity.' *Philosophical Studies* 38, 1–30.
- Davies M & Stoljar D (eds.) (2004). *Philosophical Studies* 118(1–2) (special issue on two-dimensionalism).

- Evans G (1979). 'Reference and contingency.' *The Monist* 62, 161–189.
- Kaplan D (1989). 'Demonstratives.' In Almog J, Wettstein H & Perry J (eds.) *Themes from Kaplan*. New York: Oxford University Press. 238–271.
- Kripke S (1980). *Naming and necessity*. Cambridge: Harvard University Press.
- Jackson F (1998). *From metaphysics to ethics: a defense of conceptual analysis*. New York: Oxford University Press.
- Macia J & Garcia-Carpintero M (eds.) (forthcoming). *Two-dimensional semantics*. New York: Oxford University Press.
- Segerberg K (1973). 'Two-dimensional modal logic.' *Journal of Philosophical Logic* 2, 77–96.
- Soames S (2005). *Reference and description*. Princeton, NJ: Princeton University Press.
- Stalnaker R C (1978). 'Assertion.' In Cole P (ed.) *Syntax and semantics, vol. 9: pragmatics*. New York: Academic Press. 315–322.
- Stalnaker R C (2004). 'Assertion revisited: on the interpretation of two-dimensional modal semantics.' *Philosophical Studies* 118(1–2), 299–322.
- van Fraassen B (1977). 'The only necessity is verbal necessity.' *Journal of Philosophy* 74, 71–85.

Type versus Token

L Wetzel, Georgetown University, Washington, D.C., USA

© 2006 Elsevier Ltd. All rights reserved.

The Distinction

The distinction between a 'type' of thing and a 'token' of that type of thing, first described in those terms by C. S. Peirce, is an ontological one, akin to the universal/particular distinction (but arguably different from it; see below). Peirce pointed out that English, unlike German, has only one definite article, the word 'the'; this is the word type. Yet pick up any book in English and quite a few words on a particular page are apt to be 'the's; these are word tokens. Word tokens are concrete; they are spatiotemporal particulars such as inscriptions composed of ink, raised dots, light pixels, or utterances of phonetic sounds, or smoke signals, dots and dashes, etc. (Although not every sequence of three dots followed by three dashes followed by three dots, e.g., is a token of the Morse code distress signal; it may need to have been produced with a certain intention and in accordance with certain conventions. *see Speech*

Acts.) Each token has a unique spatiotemporal location, unlike a type, which is arguably an abstract object and does not.

In other words, the word 'word' is ambiguous. When we read that there are one million words in the *O.E.D.*, that Shakespeare's vocabulary consisted of 30 000 words, or that a two-year old should be able to say 20 words, we are to count word **types**. If the printer warns us to replace the ink cartridge every 100 000 words, we are to count word **tokens**. Different counting procedures point to different criteria of identity associated with the word 'word.' The same may be said of letters, phonemes, sentences, and most other linguistic items. There are only 26 letters in the English alphabet, yet there are more than 26 letters in (your concrete token of) this sentence. The former are types, the latter tokens. There are 44 phonemes in English, yet millions of phonemes uttered every day. The numbers reverse when it comes to sentences, for linguists assure us that there are many more, perhaps infinitely many more, sentence types than token sentences produced. The type-token distinction explains and is motivated by such considerations. It helps us to avoid fallacious inferences such as that if Shakespeare knew 30 000 words, then the

total number of words in my copy of Shakespeare's plays cannot exceed 30 000. (For more about words types, see Wetzel (2000).)

Its Usefulness

The type-token distinction is important to other branches of philosophy besides philosophy of language and logic. In philosophy of mind, it yields two significantly different versions of the identity theory: one identifies types of mental events with types of physical events (suggesting that the best way to understand mental activity is through neurology); the other merely says that every mental event is some physical event or other (but not necessarily a biological physical event, leaving the door open to an understanding of mental activity in terms of, say, programming) (see Block, 1980). In aesthetics, it is customary to distinguish works of art such as Mozart's Prague symphony (a type) from its many actual performances (tokens) (and also from its many interpretations, recordings, playings of recording, etc.) (see Wollheim, 1968; Wolterstorff, 1975). In ethics, actions are said to be right/wrong, but there is a dispute as to whether there are general principles that prescribe which types of actions are right/wrong or no general principles and only action tokens that are right/wrong.

Outside of philosophy, type talk is ubiquitous. That is, in scientific and everyday discourse we often speak in ways that apparently refer to types. When, for example, we read that the mountain lion disappeared from Iowa in 1867, but now is making a comeback there in the suburbs, we know no particular cat disappeared; rather, a type of cat, a species, did so. Similarly, to say the ivory-billed woodpecker is extinct, or that the banded bog skimmer is rare is not to be referring to one particular organism. The first gene that scientists found linked to an ordinary human personality trait (novelty seeking) is obviously a type of gene. These are examples of sentences where a **singular term** apparently refers to a type. But we also **quantify** over types with great regularity, as when a study claims that of 20 481 species examined, two-thirds were secure, 7% were critically imperiled, and 15% were vulnerable. Often we do both, as when it was claimed (in the 60s) that there are 30 particles, yet all but the electron, neutrino, photon, graviton, and proton are unstable. As with art (above), so with artifacts: when we read that the personal computer is more than 30 years old and is only now beginning to reveal its true potential, we do not think there is a particular dusty 30-year-old PC that is somehow improving. The chess move, we are told, of accepting the Queen's Gambit with 2...dc has been known since

1512, but Black must be careful in this opening – the pawn snatch is too risky. For a comprehensive treatment of reference to the many sorts of abstract objects we make in discourse, see Asher (1993).

Universals

The type-token distinction is akin to the universal-particular distinction, but there is some question as to whether types ought to be classified with the classic property examples of universals such as *being white*. If the hallmark of a universal is a capacity to have more than one instance, then types are universals. Types are generally also said to be abstract, just as universals are often said to be, in contrast to their tokens, which are concrete. However, if the hallmark of a universal is to answer to a predicate or open sentence the way *being white* answers to 'is white,' then types do not resemble universals, as they answer to singular terms (and quantifiers). This is amply illustrated in the preceding paragraph. That is to say, types seem to be *objects*, like numbers and sets, rather than properties or concepts; it's just that they are not particular concrete objects. If, then, we follow Frege (1977) in classifying all objects as complete and saturated and referred to by singular terms, and all properties as incomplete and unsaturated and referred to by predicates, then types would not fall into the same category as the classic examples of universals such as *being white*.

A Related Distinction

The type-token distinction may seem to license or encourage the following fallacious inference: This encyclopedia entry consists of (about) 2,000 words. Word types, or word tokens? It can't be word types, since my editor assures me that each and every 'the' counts as another word. Therefore it consists of (about) 2,000 word tokens. But then the encyclopedia entry itself would be a token – which it is not. (It is not the one in the Library of Congress in Washington, for example, as opposed to the one in the Bibliothèque Nationale in Paris.) The entry itself has many tokens, but is itself a type. As Simons (1982) points out, since it is abstract, it cannot be composed of concrete word tokens. So what are we counting when we count 2000 words? **Occurrences** of word types (see Wetzel (1993), for an account). That is to say, the word 'the' occurs quite a few times in this encyclopedia entry. The letter 'x' occurs three times in the formula '(∃x) (Gx & Hx)'. Now this may seem impossible; how can one and the same thing occur more than once in something without there being two tokens of it? See Armstrong (1986) and Lewis

(1986a), (1986b) for a debate on this point, and Wetzel (2006) for a rebuttal of Lewis. However, consider a sequence. The number 1 – the very same number – occurs several times in the sequence $\langle 0, 1, 0, 1, 0, 1 \rangle$. Even a concrete object can occur more than once. The same person occurs more than once in the sequence of New Jersey million dollar lottery winners, remarkably enough. The moral is that not every occurrence of a type is a token of it.

Do Types Exist?

Obviously, if types are universals then whether they exist is a matter of debate – a debate that goes back at least to Plato. Modern Platonists say they exist as transcendent abstract objects with no space-time location; see Bromberger (1989), Hale (1990), Katz (1981), and Wetzel (2002). Aristotelian realists say they exist as repeatable objects with numerous space-time locations; see Armstrong (1978), Wollheim (1968), and perhaps Zemach (1992) – although he calls himself a nominalist. Nominalists say they don't exist; see Goodman and Quine (1947), Quine (1953), Goodman (1977), and Bromberger (1992). With respect to universals, there is a third position, conceptualism, which tries to make do with mental objects such as ideas, or representations, in lieu of abstract ones, but since ideas also come in types and tokens, conceptualism only postpones the question – although see Szabo (1999). Some nominalists believe that if types are abstract objects, then, as is true for all abstract objects, we cannot have knowledge of them, because we cannot causally interact with them. Others take issue with a traditional argument for universals, according to which a red sunset and a red rose have something in common, and this something can only be the property of *being red*, so properties exist. Quine (1953), for example, objects that 'the rose is red because the rose partakes of redness' is uninformative – we are no better off in terms of explanatory power with such extra objects as *redness* than we are without them; perhaps a rose's being red and a sunset's being red are just brute facts.

Since Quine's attack on the traditional argument convinced many that that argument fails, the debate has shifted away from predicates. A realist nowadays is better off pointing to the fact, evidenced in the first four paragraphs above, that we talk as if there are types (even without appealing to predicates). That is, we frequently use singular terms for types, and we quantify over types in our theories. Quine (1953) has rightly stressed that we are ontologically committed to that over which we quantify. (For these reasons, although he rejected *redness*, Quine (1987: 217) held that expression types such as the word 'red' exist.) As

we saw, Frege emphasized that singular term usage is in indicator of **objecthood**. Since at least on the face of it we are committed to types in many fields of inquiry, it is incumbent upon the nominalist to 'analyze them away.' (Or to maintain that all theories that appear to refer to types are false – but this is a radical approach, as it comprises nearly all theories.)

The usual nominalist approach is to maintain that the surface grammar of type talk is misleading, that talk of types is just a harmless way of talking about tokens (see Bromberger, 1992). To say 'the horse is a mammal' is just to say 'all horses are mammals,' for example. The idea is to 'analyze away' apparent references to types by offering translations that are apparently type-free and otherwise nominalistically acceptable (see Goodman and Quine (1947) and Goodman (1977) for optimism). The problem is how to do this for each and every type reference (see Wetzel (2000) for pessimism). Sometimes the translation is obvious as with the horse example, where the predicate 'is a mammal' is true of the type and true of all the tokens too. But not all predicates are so cooperative. The grizzly bear is ferocious, but not every grizzly is ferocious. The word 'Gluck' is guttural, but its token inscriptions are not. For that matter, the written tokens have nothing observable in common with spoken tokens (if there is no type). Collective properties pose even more of a challenge. 'The ivory-billed woodpecker is extinct' is perhaps approximately equivalent to 'there are no more ivory-billed woodpeckers, although there used to be interbreeding populations of them,' which is arguably nominalistic. But it is very difficult to find a nominalistic paraphrase for 'Old Glory had 28 stars in 1846 but now has 50.' And it is nigh-on impossible to find a paraphrase for the claim mentioned above, that 'Of 20 481 species examined, two-thirds were secure, 7% were critically imperiled, and 15% were vulnerable.'

But the clincher for this nominalistic proposal (that all type-talk can be analyzed away in favor of talk of tokens) is that it presupposes that all types have tokens. Whether all types have tokens is a matter of some debate. Peirce, for example claimed it, but Chomsky (1957) denied it. It would violate laws of syntax that together have the consequence that there are more sentences than have been or will be instantiated.

See also: Linguistic Reality; Speech Acts.

Bibliography

Armstrong D (1978). *A theory of universals*. Cambridge: Cambridge University Press.

- Armstrong D (1986). 'In defense of structural universals.' *Australasian Journal of Philosophy* 64, 85–88.
- Asher N (1993). *Reference to abstract objects in discourse*. The Netherlands: Kluwer.
- Block N (ed.) (1980). *Readings in philosophy of psychology*, 1. Cambridge, MA: Harvard University Press.
- Bromberger S (1989). 'Types and tokens in linguistics.' In George A (ed.) *Reflections on Chomsky*. Oxford: Basil Blackwell. 58–89.
- Bromberger S (1992). 'The ontology of phonology.' In his *On what we know we don't know*. Chicago: University of Chicago Press. 209–228.
- Chomsky N (1957). *Syntactic structures*. The Hague: Mouton & Co.
- Collinge N E (ed.) (1990). *An encyclopedia of language*. London: Routledge.
- Davies S (2001). *Musical works & performances: a philosophical exploration*. Oxford: Clarendon Press.
- Frege G (1977). 'On concept and object.' In Geach P & Black M (eds.) *Translations from the philosophical writings of Gottlob Frege*. Oxford: Basil Blackwell.
- Goodman N (1977). *Structure of appearance* (3rd edn.). Dordrecht, Holland: Reidel.
- Goodman N & Quine W V (1947). 'Steps toward a constructive nominalism.' *Journal of Symbolic Logic* 12, 105–122. [Reprinted 1972 in *Problems and Projects*, Indianapolis: Bobbs-Merrill. 173–198.]
- Hale B (1987). *Abstract objects*. Oxford and New York: Basil Blackwell.
- Hutton C (1990). *Abstraction & instance: the type-token relation in linguistic theory*. Oxford: Pergamon Press 1–19.
- Katz J J (1981). *Languages and other abstract objects*. Totawa, N. J.: Rowman and Littlefield.
- Lewis D (1986a). 'Against structural universals.' *Australasian Journal of Philosophy* 64, 25–46.
- Lewis D (1986b). 'Comment on Armstrong and Forrest.' *Australasian Journal of Philosophy* 64, 92–93.
- Peirce C S (1931–1958). *Collected papers of Charles Sanders Peirce*. In Hartshorne & Weiss (eds.). Cambridge: Harvard University Press.
- Quine W V (1953). 'On what there is.' In *From a logical point of view*. Cambridge, MA: Harvard University Press. 1–19.
- Quine W V (1987). *Quiddities: an intermittently philosophical dictionary*. Cambridge, MA: Harvard University Press. 216–219.
- Simons P (1982). 'Token resistance.' *Analysis* 42(4), 195–203.
- Szabo Z (1999). 'Expressions and their representations.' *The Philosophical Quarterly* 49, 145–163.
- Wetzel L (1993). 'What are occurrences of expressions?' *Journal of Philosophical Logic* 22, 215–220.
- Wetzel L (2000). 'The trouble with nominalism.' *Philosophical Studies*, 98, 361–370.
- Wetzel L (2002). 'On types and words.' *Journal of Philosophical Research* 27, 239–265.
- Wetzel L (2006). *Types and tokens: an essay on universals*. Cambridge, MA.: MIT. Press.
- Wollheim R (1968). *Art and its objects*. New York: Harper and Row.
- Wolterstorff N (1975). 'Toward an ontology of art works.' *Nous* 9, 115–142.
- Zemach E (1992). *Types: essays in metaphysics*. Leiden, The Netherlands: E. J. Brill.
-



Use Theories of Meaning

G Marti, ICREA & Universitat de Barcelona,
Barcelona, Spain

© 2006 Elsevier Ltd. All rights reserved.

Traditional theories of meaning of the kind proposed by Frege, Russell, and the early Wittgenstein take to heart the idea that language is a system of symbols whose essential role is to **state** or to **represent** the goings on of the world and the mind. From the point of view of a traditional, or representationalist, approach to meaning, the fundamental area of inquiry about language and its significance has to do with the connection between linguistic items and the things they stand for or represent, things or facts. That is why two of the main focal points of traditional theories of meaning are the theory of reference – the exploration of the bond that ties expressions in a language to things in the world – and the theory of propositions – the discussion of the form and the constitution of what is expressed by utterances of sentences and their role in the determination of truth or falsity. So-called use theories of meaning can be best seen as reactions to the fundamental tenets of traditional theories: whereas traditional theories focus on what language represents and how it represents it, use theories search for the key to meaning in actual usage and linguistic practice.

The first strong appeal to use in the theory of meaning appeared in print in 1950, in P. F. Strawson's article 'On Referring.' Strawson's article is meant to be primarily a critical response to the analysis of sentences containing definite descriptions proposed by Bertrand Russell in his seminal paper 'On Denoting.' The merits or demerits of Strawson's specific objections to Russell's theory of descriptions will not be addressed here. The important point for our purposes is rather the positive outlook on meaning that Strawson's remarks suggest. In criticizing Russell, Strawson (1971: 9) contends that "to talk about the meaning of an expression or sentence [is to talk about] the rules, habits, conventions governing its correct use, on all occasions to refer or to assert". Although rather programmatic in character,

Strawson's remarks point towards a conception of the significance of linguistic expressions that departs radically from the Frege–Russell–early Wittgenstein approach for, according to Strawson, neither the things that terms stand for, as Russell and Wittgenstein would have it, nor the conceptual material associated with expressions – the sense – that in turn determines what those expressions stand for, as Frege would have it, constitute appropriate answers to the question 'What is the meaning of X?' On the contrary: "to give the meaning of an expression is to give general directions for its use" (Strawson, 1971: 9).

Traditional theories of meaning, with their emphasis on reference and propositions, leave out of the realm of semantic inquiry a host of expressions whose significance cannot be doubted. For what does 'hello' refer to? What proposition do we express, what fragment of the world do we represent, when we say 'hello'? If we do express a proposition, under which conditions is it true or false? Unlike traditional theories, use theories of meaning would explain the meaning of 'hello' and similar expressions by appeal to the rules and conventions that indicate their appropriate use. But it is not only expressions such as 'hello,' 'ouch,' or 'pardon' that, according to use theorists, are left behind by traditional approaches. J. L. Austin argued that this is so for a class of sentences, which he characterized as 'performative,' that look grammatically like any other subject-predicate or subject-verb-object sentence that we use to represent a fragment of the world, such as 'It is raining' or 'The train is late.' We use the latter to tell the way things are and consequently when we utter them we say something true or something false. Performative utterances, on the other hand, do not state or represent the way things are. When we utter a performative, Austin argues, we certainly say something significant, yet it is neither true nor false. The uses of sentences that Austin is thinking of are utterances of, for instance, 'I name this ship *Queen Elizabeth*,' 'I do [take this man to be my husband],' 'I promise to be at the party,' 'I apologize.' According to Austin what all these sentences have in common is that by uttering one of them we do not report or describe

a state of affairs, an event or an action. Uttering one of those sentences is actually **performing** an action. Uttering a performative is a speech act. By saying 'I do,' as Austin (1961: 235) puts it, "I am not reporting on a marriage, I am indulging in it". Because utterances of performatives do not function like statements, the question of truth or falsity does not arise with regard to them. However, there are, Austin points out, ways in which performatives can succeed or fail. For instance, if the utterer does not have the authority to give a name to a ship, her utterance of 'I name this ship *Queen Elizabeth*' does not succeed in performing the intended action. For an utterance of a performative to be satisfactory some felicity conditions must be in place.

It is tempting to think that Austin is highlighting a phenomenon circumscribed to a well-delimited class of utterances, and that for the vast majority of natural language sentences the traditional representational picture applies smoothly. That is not so: what Austin is proposing is not just a specific treatment of a peculiar phenomenon but rather a different conception of meaningfulness. On Austin's view, statements, as much as performatives, are essentially speech acts and are therefore subject to similar conditions of adequacy and success. Suppose we utter a statement such as 'John's children are all very polite' and John has no children. According to Austin, the situation as regards that statement parallels a performative utterance such as 'I promise I will sell you this piece of land' when the piece of land in question does not exist. In the latter case we would say that the sale is void; and in the former case we should say that the statement is void also. Both performatives and statements are subject to the question 'Is it in order?' If the answer is negative, the performance fails. And if the answer is positive, then both performatives and statements are subject to further questions of felicity. Those questions may take different forms depending on the type of speech act: if it is a warning, the question is whether it was justified; if it is a piece of advice, the question is whether it was sound; if it is a statement, the question is whether it was true; in every case, although in different forms, those questions "can only be decided by considering how the content ... is related in some way to fact ... we do require to assess at least a great many performative utterances in a general dimension of correspondence with fact" (Austin, 1961: 250). In this way Austin (1961: 251) takes statements "off their pedestal" and offers a uniform picture of meaning that appeals essentially to our usage of expressions to do things.

Among use theorists we cannot forget Ludwig Wittgenstein, the coiner of the motto 'meaning is

use.' In the first part of the *Philosophical Investigations*, completed in 1945 although not published until 1953, Wittgenstein reacts strongly against the traditional conception of meaning. In fact, Wittgenstein is reacting against his own earlier views presented in the *Tractatus Logico-Philosophicus*. Wittgenstein's account of meaning in the *Tractatus*, with its characterization of the proposition as a picture of a fragment of the world, is the paramount example of a traditional representationalist approach. In the *Investigations*, by contrast, Wittgenstein rejects the explanatory project of Frege and Russell: neither the Fregean senses that are supposed to determine the objects expressions refer to, nor the things that on Russell's (and the *Tractatus*'s) approach are directly and conventionally associated with signs, give life and significance to language: only its use gives meaning to a sign.

To understand the kind of practice we engage in when we use a language as members of a speakers' community, Wittgenstein appeals to the metaphor of playing a game. For Wittgenstein, learning a language, like playing tennis, consists in becoming competent at a rule-governed practice. It does not consist in explicitly learning the rules, having them, so to speak, written in one's mind; being able to play tennis, or being able to speak and understand a language consists in being proficient at doing something according to the rules.

Different use-oriented approaches to meaning criticize the traditionalist stance for disregarding the fact that language is a tool that we use to do things, that speaking and understanding a language is a matter of engaging in a practice, and that, consequently, the key to meaning is to be found in the way language users employ language. But how radical the departure from the traditional stance is depends on how the motto 'meaning is use' is interpreted. On the one hand it may be interpreted as a claim about what gives expressions the meaning they have. The appeal to use is then a reminder that languages are social institutions and that it is by virtue of usage that expressions are connected to their meanings. From this point of view, it is not inappropriate to say, for instance, that by virtue of its use, the word 'dog' means a certain concept or that its meaning consists in naming a species. So interpreted, the claim that meaning is use is fundamentally presemantic, it is not a claim about what constitutes the meaning of expressions. So conceived, use theories of meaning are not opposed in essence to traditional theories, although by stressing the meaning-conferring role of use they expand the horizons of the traditional stance, for they make room in semantic theory for expressions whose meaning cannot be cashed out in terms of what and how

they represent objects or states of affairs, and they do not disregard aspects of meaning that are not truth conditional.

A more radical way of interpreting the claim that meaning is use is as a semantic claim, i.e., a claim about what constitutes the meaning of linguistic expressions. From this point of view there is nothing over and above the way an expression is used that can qualify as its meaning. Use is not just what makes an expression have a meaning; it is **all there is** to meaning. Some varieties of deflationism take this stance. Thus, for instance, in Horwich (1998: 6) we read: "The meaning property of a word reduces [to] . . . the property that every use of the word is explained in terms of the fact that we accept certain specified sentences containing it."

The radical interpretation of the claim that meaning is use faces a number of objections. Here I will focus on only two general challenges (for discussion, see Horwich, 1998). First, it may be argued that the idea that use determines the meaning of an expression puts the cart before the horse. Intuitively there is a distinction to be drawn between correct or incorrect usage. No matter how pervasive the use of 'irregardless' is, it is an incorrect expression. The very idea of incorrect, but extended, use seems to entail that it is because expressions do have a meaning, over and above the way in which they are used, that we can talk about correct or incorrect usage. Meaning determines (correct) use and not vice versa, so it seems that the claim that meaning is constituted by use has difficulties accounting for the normative aspect of meaning.

Second, the idea that all there is to the meaning of an expression is its use does not leave room for what appears to be a legitimate possibility: a speaker may be competent in the use of an expression, she may know the situations in which it is appropriate to use it and how to react to uses of it, and yet she may not know the meaning of the expression in question. The possibility is, in fact, less far fetched than it appears. Consider the case of Helen Keller: blind and deaf from an early age, she explains in her autobiography

how she and her family had developed a rather good system of symbols to communicate their needs and wishes: when she wanted bread, she made a sign, when she wanted ice cream, she made another sign. Her mother had ways to tell Helen what she needed, and Helen would go and get it for her: Helen had mastered the use of a system of symbols. Nevertheless, when she was seven her teacher, Miss Sullivan, once spelled on Helen's hand the sign for water while Helen felt water with her other hand, and that was the moment that she describes as "learning the key to all language." What could she possibly have discovered that she didn't know before? It surely was not how to use the sign 'water' but rather, and quite simply, that 'water' stands for water. Now, it may be argued that learning that 'water' can be used to refer to water is indeed learning something new about the use of 'water.' But then it would appear that even from the point of view of use theories of meaning we need to be sensitive to the fact that expressions represent things and that it is the relation between words and things that makes it possible for us to talk about the world.

See also: Deflationism; Normativity; Speech Acts.

Bibliography

- Austin J L (1961). 'Performative utterances.' *Philosophical Papers*. Oxford: Oxford University Press.
- Austin J L (1962). *How to do things with words. The William James Lectures delivered at Harvard University in 1955*. Cambridge, MA: Harvard University Press.
- Horwich P (1962). *Meaning*. Oxford: Clarendon Press.
- Keller H (1962). *The story of my life*. New York: Penguin.
- Strawson P F (1962). 'On referring.' In *Logico-Linguistic Papers*. London: Methuen. Originally published in *Mind*, vol. LIX, 1950.
- Wittgenstein L (1962). *Philosophical investigations*. New York: Macmillan.
- Wittgenstein L (1962). *Tractatus logico-philosophicus*. London: Routledge and Kegan Paul (first published in English in 1922).

Use versus Mention

C Spencer, Howard University, Washington, DC, USA

© 2006 Elsevier Ltd. All rights reserved.

Generally, the distinction between using something and mentioning it is completely straightforward. For

instance, there is all the difference in the world between mentioning, or talking about, a lawn mower, and using it to mow the lawn. Yet the distinction can engender some confusion when it comes to mental and linguistic representations, and this is what philosophers and linguists have in mind when they

talk about the use/mention distinction. The distinction between mentioning a word, picture, or mental representation, and using it to communicate or entertain a thought is best illustrated by example. Sentence (1) uses the word 'Boston,' and sentence (2) mentions it.

- (1) Fred is from Boston.
- (2) 'Boston' has two syllables.

Philosophers customarily put a word (or complex expression) inside quotation marks, to form its 'quote name,' if they want to mention it instead of using it, as in (2). Linguists, in contrast, typically use italics, labeled bracketing trees, or a phonetic alphabet to mention an expression, as in (3a–c).

- (3a) *Boston* has two syllables.
- (3b) [_{NP} Boston] has two syllables.
- (3c) bɛs tən has two syllables.

An advantage of the philosopher's convention is that we may easily form the quote name of a quote name simply by enclosing the quote name itself in quotation marks, as in (4):

- (4) "“Boston”" is the name of a word, not a city.

We can also mention an expression without using any of these devices, as in (5) and (6).

- (5) The first line of Gray's *Elegy* states a proposition.
- (6) Samuel Clemens's pen name is derived from a riverboat driver's call.

Example (5) uses an expression, 'the first line of Gray's *Elegy*,' to mention the linguistic expression it denotes, namely, 'The curfew tolls the knell of parting day.' Similarly in (6), the expression 'Samuel Clemens's pen name' is used to mention another linguistic expression, namely 'Mark Twain.'

When we talk more systematically about a particular language, as we do in describing its syntax or semantics, we observe the related distinction between object language and metalanguage. The object language is the language under discussion, and the metalanguage is the language we use to talk about the object language. In some cases, the object language and the metalanguage are different languages, as in (7), but the object language and the metalanguage can be the very same, as in (8).

- (7) 'Le chien est sur la chaise' is true if and only if the dog is on the chair.
- (8) 'The dog is on the chair' is true if and only if the dog is on the chair.

Sentences (7) and (8) mention sentences of the object language (enclosed within quotation marks). The metalanguage, in both cases English, is the language used to state the truth condition of sentences in the object language.

Sentences that mention themselves, such as (9), can give rise to paradox.

- (9) This sentence is false.

Sentence (9) appears to have the paradoxical property of expressing a truth if and only if it expresses a falsehood. Many sentences that mention themselves are innocent in this respect, such as (10):

- (10) This sentence contains five words.

The paradox only arises for sentences that self-ascribe semantic properties such as truth or falsehood. A central project of a theory of truth is to resolve this paradox.

So far, the use/mention distinction seems easy to keep straight: How could someone confuse the name 'Boston' with the New England city itself? Yet philosophers sometimes charge one another with confusing use and mention of linguistic expressions. Such confusion typically results in mere obscurity, as it does in this passage from Leibniz, as translated by C. I. Lewis.

Two terms are the same if one can be substituted for the other without altering the truth of any statement. If we have A and B and A enters into some true proposition, and the substitution of B for A wherever it appears, results in a new proposition which is likewise true, and if this can be done for every such proposition, then A and B are said to be the same; and conversely, if A and B are the same, then they can be substituted for one another as I have said (Lewis, 1960: 291).

Richard Cartwright has suggested that the apparent use/mention confusion here makes it difficult to see exactly what Leibniz intended to say in this passage (Cartwright, 1971: 119). Perhaps Leibniz meant that A and B are the same if and only if **their names** can be intersubstituted into any statement without altering the truth value of the proposition that the statement expresses. If this was what he meant, then in some places he should have **mentioned** the names of A and B rather than **using** them, as he in fact did.

In other cases, however, use/mention confusions may ground substantive philosophical mistakes. One such mistake is the familiar confusion of objects with our ideas of them. This mistake arises particularly for philosophically problematic entities such as the number three and Beethoven's Ninth Symphony. It is obvious that they are not concrete objects, but less obvious what else they could be. Since we clearly have ideas of these objects, some have suggested that we identify the objects themselves with our ideas of them. cursory reflection reveals that this suggestion cannot be correct, since these entities have properties

that ideas could not have. Ideas, for instance, cannot be the cube root of 27 or played frequently by the Berlin Philharmonic. Plausibly, what lies behind this suggestion that objects like these are really ideas in our heads is a confusion of our mental representations of these objects with the objects themselves. Like the confusion of the word 'Boston' with the city of Boston, this is a straightforward confusion of use with mention.

See also: Limits of Language; Metalanguage versus Object Language; Truth: Theories of in Philosophy.

Bibliography

- Cartwright R C (1971). 'Identity and substitutivity.' In Munitz M K (ed.) *Identity and individuation*. New York: New York University Press. 119–133.
- Goldfarb W (2003). 'Use and mention.' In Goldfarb W *Deductive logic*. Indianapolis: Hackett.
- Lewis C I (1960). *A survey of symbolic logic*. New York: Dover.
- Martinich A P (1990). 'Introduction.' In Martinich A P (ed.) *The philosophy of language*, 2nd edn. New York: Oxford University Press.
- Quine W V O (1970). *Philosophy of logic*. Cambridge, MA: Harvard University Press.

This page intentionally left blank

V

Vagueness: Philosophical Aspects

R Keefe, University of Sheffield, Sheffield, UK

© 2006 Elsevier Ltd. All rights reserved.

Hallmarks of Vagueness

Much of language is vague, including expressions such as ‘heap,’ ‘bald,’ ‘red,’ ‘child,’ and ‘hungry.’ Although the exact criteria of vagueness are controversial, we can say that, at least in general, vague predicates have borderline cases and seem to lack sharp boundaries and well-defined extensions. A dieting man, Don, may at times be a borderline case of ‘thin’ – it is unclear whether or not he is thin. And the apparent lack of sharp boundaries to the predicate is exhibited in the fact that there does not appear to be a sudden instant during his successful diet at which Don becomes thin. Though we frequently call assertions vague if they don’t supply us with enough information (e.g., ‘I’ll visit next week’), this is a different sense of ‘vague’ from that at issue here.

Vague predicates give rise to sorites paradoxes, of which the most famous is the paradox of the heap. If you have a heap of sand, and you remove a single grain, then the result will still be a heap. But then, suppose we take a heap of sand and remove grains one by one; repeated applications of that principle imply absurdly that the solitary last grain is a heap. The vagueness of other predicates accounts for the plausibility of comparable principles that can figure in sorites paradoxes, such as ‘if someone is a child at time t , then she/he is a child 10 minutes later’ or ‘if someone of height h is tall, then so is someone one-hundredth of an inch shorter.’

Typical vague predicates also exhibit what is known as ‘higher-order vagueness.’ Not only does ‘tall’ have borderline cases at heights between the definitely tall and the definitely not-tall, but the set of such borderline cases does not seem to have sharp boundaries either. Just as it is not plausible to think there is a precise instant at which Don becomes thin, there seems to be no precise instant at which Don becomes borderline thin, or one at which he ceases to be borderline thin and becomes definitely

thin. Perhaps, then, the borderline cases of ‘thin’ themselves have borderline cases – borderline borderline cases of ‘thin.’ And, moreover, those borderline borderline cases do not seem to be sharply bounded either, suggesting a hierarchy of borderline cases corresponding to the apparent lack of sharp boundaries at all levels.

Finally, note that other elements of language besides predicates are also vague: consider, for example, ‘hurries,’ ‘quickly,’ ‘many,’ ‘very’ and ‘the nicest man.’ Analogues of the central hallmarks of vagueness could be described in relation to these kinds of expressions.

Three Philosophical Debates About Vagueness

First, is vagueness eliminable? Many have thought not. Vague predicates seem to be an unavoidable presence in our language. As Russell (1923) argues, the application of, say, ‘red’ is based on unaided sense perception, and so it surely cannot be applicable to only one of an indiscriminable pair. But this thought is enough to generate a sorites paradox by considering something that changes from red to yellow through a series of changes too small for us to detect. Similarly, it can be important to have a word, like ‘tall,’ to pick out a vague range of a measurable quantity, like height. The applicability of other useful vague predicates can depend on several dimensions of variation – whether someone counts as ‘big’ can depend on both their height and their weight – or it can depend on a range of dimensions that is itself not well-defined – such as ‘nice’ applied to people.

Second, is there ontological vagueness? It is controversial whether it is only language and thought that can properly be called vague, or whether there can be vague objects or other vague entities. Russell thought that only representations could be vague, and Evans (1978) has argued that the idea of vague objects is not coherent, but other philosophers have attempted to make sense of the idea. This issue of metaphysical vagueness is often tackled independently of the

attempt to give a theory of linguistic vagueness, with which we are concerned here.

Third, does vagueness have implications for logic? Many have thought that vagueness poses a challenge to classical logic. If Don is a borderline case of thinness, it is tempting to maintain that 'Don is thin' is neither true nor false, which is incompatible with the principle of bivalence. And the apparent lack of sharp boundaries to vague predicates is in tension with the well-defined extensions required for classical logic.

Philosophical Theories of Vagueness

A theory of vagueness faces the task of solving the sorites paradox and determining the logic and semantics of vague language.

One approach is to reject classical logic, acknowledging that borderline case predications are neither true nor false, and specifying that they have some intermediate truth-value. For example, we could adopt a three-valued logic and explain the apparent lack of sharp boundaries between the Fs and the non-Fs in terms of the range of cases between these two poles (see Tye, 1994). Or we could recognize a full range of intermediate truth-values corresponding to degrees of truth: in the region of transition from not-thin to thin, Don may count as thin to different degrees, gradually increasing from the minimum, 0, to the maximum, 1. This kind of degree theory calls for an infinite-valued logic (see Machina, 1976; for a non-standard degree theory, see Edgington, 1997).

One common objection to this kind of approach is that it merely replaces classical logic's sharp boundary between truth and falsity with other sharp boundaries between adjacent members of the new set of truth-values. So, for example, according to a three-valued logic, there will be a first instant at which Don ceases to be falsely called thin becomes borderline thin, and with a degree theory, there will similarly be a first instant at which he is thin to some degree greater than 0. Intuitively, there are no such sharp boundaries either. One way of putting the worry is that these views of vagueness cannot accommodate higher-order vagueness. Some theorists, however, (e.g., Tye) have attempted to defend a many-valued theory that is not committed to sharp transitions by appealing to the vagueness of the metalanguage in which the theory is formulated.

An alternative approach to vagueness involves accepting sharp boundaries at the first level and embracing classical logic. This epistemic view maintains that borderline case predications are either true or false – we just do not know which (see Williamson, 1994). Similarly, there is a particular instant at which

Don becomes thin – we just do not know when it is. This solves the sorites paradox by denying its main premise, that is, the principle that taking a grain away from a heap will leave you with a heap. Classical logic is preserved in its entirety. The apparent lack of sharp boundaries is merely apparent, and vagueness is a matter of ignorance. Williamson (1994) has defended such a theory and provided a detailed explanation of why we are ignorant by appealing to the idea of a margin for error. If I know that someone of height h is tall, then someone of height $h - e$ is tall, where e is less than the margin required for my belief to be reliable. It follows that I can't know that someone slightly above the boundary is tall nor, by a parallel principle, can I know that someone slightly below it is not tall; so I can't know that the boundary falls between these two close points.

The supervaluationist theory of vagueness takes a different approach to the semantics of vague language while preserving the theorems and inferences of classical logic (see Fine, 1975; Keefe, 2000). According to this theory, 'Don is thin' is true (false) if and only if it is true (false) on all ways of making it precise, i.e., with all reasonable ways of drawing the boundary to 'thin.' If Don is borderline thin, that sentence is true on some of those ways and false on others, so it is neither true nor false. Before his diet, however, Don was definitely not thin, and it counted as false that he is thin according to supervaluationism, because any boundary that ruled him as thin at that time is not an acceptable way to make 'thin' precise. To see how the law of excluded middle still holds, despite the failure of bivalence, consider the instance 'either Don is thin or he isn't' in the case where he is borderline thin. On every way of making 'thin' precise, that disjunction comes out true, even though on some of those ways it is the first disjunct that is true, and on others it is the second disjunct. The main premise of the sorites paradox comes out false, on the other hand: on every way of drawing the boundary to 'heap,' there is some grain the removal of which turns a heap into a non-heap. But 'heap' is nonetheless not sharply bounded, because there is no particular grain whose removal marks the boundary to 'heap'; different grains mark the boundary for different ways of making 'heap' precise.

Other solutions to the sorites paradox have been proposed in addition to those embedded in the influential theories just sketched. Contextualist accounts appeal to changes in context between the assessments of different predications along a sorites series (see, e.g., Graff, 2000). Now, vague terms are typically context-dependent; whether someone counts as thin or something counts as a heap can depend on the context in which it is considered. But vagueness does not

straightforwardly reduce to context-dependence, because if we fix the context, a vague term will still have borderline cases and be susceptible to a sorites paradox. Nonetheless, the contextualist approach attempts to close the gap between these phenomena with a complex and subtle treatment of context-change.

The nihilist response to the sorites paradox is one of the few that does not involve denying its main premise (see Unger, 1979). The meaning of 'heap' dictates that a heap will always survive the removal of a single grain, but the paradoxical argument does not take hold because there are no heaps. Run the case by adding grains one at a time to a single grain (adding a single grain to something that isn't a heap doesn't create a heap) and you have a sound argument with a true conclusion. This nihilist position thus renders our vague predicates empty; there are no tall people, thin people, or red things either.

To summarize, vague predicates typically have borderline cases, apparently lack sharp boundaries, and are susceptible to sorites paradoxes. For example 'is a ripe fruit' has borderline cases – a banana that is not definitely ripe and not definitely not ripe – and there is not typically a sharp point at which a fruit becomes ripe. And we can construct a sorites paradox with the notion using a premise such as 'if a fruit is not ripe at some instant, then it is not ripe one second later.' Vagueness, recall, is not merely an eliminable defect of our language; no precise replacement for 'ripe' would be usable in practice. And it poses a challenge to principles of classical logic, increasing the urgency of providing a theory of vagueness. Theories adopting a multi-valued logic will assign some intermediate truth-value between truth and falsity to the statement that our borderline-ripe banana is ripe. An epistemicist theory will maintain that such a statement is either true or false, but we just don't know which. And the supervaluationist will claim that it is neither true nor false, because it is true on some ways of making 'ripe' precise and false on others. Each of these views solves the sorites paradox by denying

that the main premise is completely true, though no matter what they do to explain this, it is hard to dismiss the intuitiveness of a premise like 'if a fruit is not ripe at some instant, then it is not ripe one second later.'

Bibliography

- Beall J C (ed.) (2003). *Liars and heaps: new essays on paradox*. Oxford: Oxford University Press.
- Edgington D (1997). 'Vagueness by degrees.' In Keefe R & Smith P (eds.). 294–316.
- Evans G (1978). 'Can there be vague objects?' *Analysis* 38, 208. In Keefe R & Smith P (eds.). 317.
- Fine K (1975). 'Vagueness, truth and logic.' *Synthese* 30, 265–300. In Keefe R & Smith P (eds.). 117–150.
- Graff D (2000). 'Shifting sands: an interest-relative theory of vagueness.' *Philosophical Topics* 28, 45–81.
- Horgan T (1994). 'Robust vagueness and the forced-march sorites paradox.' In Tomberlin J E (ed.) *Philosophical perspectives 8: logic and language*. Atascadero, CA: Ridgeview.
- Keefe R (2000). *Theories of vagueness*. Cambridge: Cambridge University Press.
- Keefe R & Smith P (eds.) (1997). *Vagueness: a reader*. Cambridge, MA: MIT Press.
- Machina K F (1976). 'Truth, belief, and vagueness.' *Journal of Philosophical Logic* 5, 47–78. In Keefe R & Smith P (eds.). 174–203.
- Russell B (1923). 'Vagueness.' *Australasian Journal of Philosophy and Psychology* 1, 84–92. In Keefe R & Smith P (eds.). 61–68.
- Sainsbury R M (1990). 'Concepts without boundaries.' In Keefe R & Smith P (eds.). *Inaugural lecture published by the King's College London Department of Philosophy*. 251–264.
- Tye M (1994). 'Sorites paradoxes and the semantics of vagueness.' Orig. in Tomberlin J E (ed.). *Philosophical perspectives, 8: logic and language*. Atascadero, CA: Ridgeview. In Keefe R & Smith P (eds.) (with omissions), 281–293.
- Unger P (1979). 'There are no ordinary things.' *Synthese* 41, 117–154.
- Williamson T (1994). *Vagueness*. London: Routledge.

Verificationism

M Beaney, University of York, York, UK

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Verificationism is the view that the meaning of a (synthetic or empirical) statement is given by its method of verification. A sentence, as used on a given occasion to make a (synthetic or empirical) statement, has meaning if and only if its truth or falsity can – in principle – be determined by experience. Verificationism was the central doctrine of logical positivism (also called ‘logical empiricism’), a movement that originated in the work of the Vienna circle in the early 1930s and received its classic statement in A. J. Ayer’s *Language, truth and logic* (1936). Although subject to devastating criticism in the 1940s and 1950s, the motivation behind verificationism has continued to influence philosophers ever since, most notably, in the work of W. V. O. Quine and Michael Dummett. Indeed, the basic positivist impulse – to reject anything that is not grounded in sensory experience – goes back at least to David Hume, and has been a significant feature of the philosophical landscape throughout the modern era. Hume’s famous words at the very end of his *Enquiry concerning human understanding* are often taken as the definitive statement of the underlying positivist view:

If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.

The repudiation of metaphysics was characteristic of logical positivism, too, and this was rooted in the doctrine of verificationism.

The Analytic/Synthetic Distinction

Central to logical positivism was the distinction between analytic and synthetic statements – corresponding to Hume’s distinction (implicit in the passage just quoted) between relations of ideas and matters of fact. According to the logical positivists, a statement is analytic if and only if its truth or falsity is determined solely by the meaning of its constituent terms. ‘All bachelors are unmarried men,’ for example, was seen as true in virtue of the meaning of the term ‘bachelor.’ Analytic truths were regarded as both necessary and *a priori*. Their necessity was seen to lie

in their tautological nature, a view that the logical positivists took from Ludwig Wittgenstein. Since, on their account, analytic truths could be judged to be true merely by knowing their meaning – without needing to consult the world in any way – they were also regarded as *a priori* truths.

A statement was seen as synthetic, on the other hand, if it is not analytic, and synthetic truths were regarded as both contingent and *a posteriori*. It was in explaining synthetic statements that the doctrine of verificationism was formulated. According to this doctrine, the meaning of such a statement lies in its method of verification. A statement is meaningful if and only if its truth or falsity can in principle be determined by experience, that is, be derived in some specified way from the truth of one or more observation statements – statements that record the direct result of an observation. This characterization immediately raises two questions. What exactly is the relationship between the statement whose truth-value is to be determined and the observation statements? And what exactly is an observation statement?

Observation Statements

Let us take the latter question first. There was much debate among the members of the Vienna circle in the 1920s and 1930s about the foundations of empirical knowledge and the correct form that observation statements – or ‘protocol sentences,’ as they were called – should take. In his *Aufbau* of 1928, influenced by Bertrand Russell’s appeal to sense data in *Our knowledge of the external world* (1914), Rudolf Carnap sketched a ‘reconstruction’ of our empirical knowledge on a phenomenalistic base. But he also offered the possibility of a physicalistic reduction, reflecting his rejection of the idea that there is any privileged ontology. (Cf. Beaney, 2004: § 4.) Despite this, however, a physicalist language soon became seen as the preferred form in which the protocol sentences were to be expressed, although debate raged about the relations between the physical and the psychological, and indeed, about the very idea of a ‘foundation.’ (See, e.g., Carnap, 1932/1933; Neurath, 1932/1933; Schlick, 1934; Ayer, 1936/1937. For an excellent account of the debate, see Uebel, 1992.)

In A. J. Ayer’s work, however, influenced by Russell and the British empiricists, observation statements were construed as statements about sense data, and material objects – and indeed, other minds – were seen as logical constructions out of them. The project of logical construction was all the rage among analytic

philosophers in the early 1930s (see especially Wisdom, 1931–1933), but it gradually became clear just what difficulties it faced. (For a critique, see, e.g., Urmson, 1962, ch. 10, and more recently, Soames, 2003, I, ch. 7.) More fundamentally, the whole appeal to sense data proved problematic, as the very idea of sense data came under increasing fire in the years that followed. (See especially Austin, 1962, and the essays in section II of Swartz, 1965.)

Strong Verification

Problems arose, then, about the nature of observation statements. But even if we leave these aside, and assume that there is a legitimate class of observation statements, such as expressed by sentences of the form ‘The *A* in front of me is a *B*,’ there remains the question of specifying the relationship between the synthetic statement whose truth-value is to be determined and the observation statements. The relationship was seen as one of verification. But what is it to verify a statement? A distinction came to be drawn between ‘strong’ and ‘weak’ verification. In the strong sense, a statement is verifiable if and only if, as Ayer put it, “its truth could be conclusively established in experience” (1946: 12). More precisely, the notion might be defined as follows: a statement *S* is strongly verifiable if and only if there is some set of observation statements which logically entail *S*. The obvious problem with this, however, is that universal generalizations cannot be strongly verified, since no (finite) set of observation statements, ‘*A*₁ is *B*,’ ‘*A*₂ is *B*,’ . . . ‘*A*_{*n*} is *B*,’ entails ‘All *As* are *B*.’ Even if all the *As* that we have so far experienced have been *B*, there remains the possibility that the next *A* we experience will not be a *B*. Yet many universal generalizations are clearly meaningful, so we cannot accept strong verifiability as the criterion for meaningfulness.

Strong Verification and Strong Falsification

This might suggest that we should talk not of verifiability but of falsifiability (as Karl Popper was to urge). A statement *S* is strongly falsifiable, we might then say, if and only if there is some set of observation statements which logically entail not-*S*. Saying that a statement is meaningful if and only if it is strongly falsifiable legitimizes universal generalizations, since it only takes one observation statement as a counterexample to falsify a universal generalization. However, we are then faced with the corresponding problem of existential statements, such as ‘There is at least one *A* that is not a *B*’ (the contradictory of ‘All *As* are *B*’). Finding an *A* that is not a *B* may verify this statement, but no observation statement could falsify

it. Finding an *A* that is a *B* neither verifies nor falsifies it. To falsify it we need to be sure that *all As* are *B*, but as we have just seen, this is not entailed by any set of observation statements that we may have made up to now.

The obvious response to both these problems is to say that a statement is meaningful if and only if it is *either* strongly verifiable *or* strongly falsifiable. But this, too, faces problems. First, there are statements involving more complex or different kinds of quantification, such as those expressed by sentences of the form ‘For every *A*, there is a *B* to which it is *R*-related’ and ‘Most *As* are *B*.’ Although many of these statements are meaningful, they, too, are neither strongly verifiable nor strongly falsifiable. Second, there are all sorts of statements that scientists make, positing the existence of unobservable things such as electrons, charge, and gravity, whose truth-value cannot be simply deduced from any set of observation statements. Rather, their justification lies in the explanation that they offer of observable events. Talk of strong verification and falsification does not do justice to a fundamental method of science – inference to the best explanation (cf. Soames, 2003, I: 280–282).

Weak Verification

The only answer is to retreat to a weaker notion of verification. In the weak sense, a statement is verifiable (and hence meaningful) if and only if, as Ayer put it, “it is possible for experience to render it probable” (1946: 12). Instead of seeing the important relation as deducibility *from* observation statements *to* the statement to be verified, the focus now is on deducibility *from* the statement to be verified *to* observation statements (which, if true, lend it empirical support). Ayer’s initial attempt at a definition may be formulated as follows: a synthetic statement *S* is weakly verifiable (and hence meaningful) if and only if *S*, either by itself or in conjunction with certain other premises, logically entails some observation statement *O* that is not entailed by those other premises alone. (Cf. Ayer, 1946: 15; Soames, 2003, I: 283.) On this definition universal generalizations, existential statements, more complex quantified statements, as well as hypothesized scientific explanations of observable events, all come out as meaningful, since observation statements can be deduced from them.

However, as it stands, this definition has the result that *all* statements come out as meaningful. For consider taking the single additional premise, ‘If *S* then *O*.’ From ‘*S*’ and ‘If *S* then *O*,’ ‘*O*’ clearly follows, without following from ‘If *S* then *O*’ alone. So the notion of verification formulated here is far too weak: the definition is satisfied by any statement whatever.

In response, Ayer distinguished between direct and indirect verifiability, restricted what ‘other premises’ are allowed, and formulated a more complex definition. (Cf. Ayer, 1946: 15–18.) But this too has been found to generate similar problems, with suitable choice of additional premises; and alternative formulations have also been found to be open to objections. (Cf. Church, 1949; Hempel, 1950; Soames, 2003, I: 284–291. But for a recent attempt at reformulation, see Wright, 1993, ch. 10.)

The Influence of Verificationism

If an adequate notion of verification cannot be formulated, then it undermines its role in a critique of metaphysics. One of its purposes had been to distinguish legitimate scientific statements from meaningless metaphysical statements – exemplified for the logical positivists in Heidegger’s famous remark that ‘Nothingness itself nothings’ (*Das Nichts selbst nichtet*) (cf. Carnap, 1932: § 5; Friedman, 2000: ch. 2). But it might well be argued that claims about the doctrine of verification itself are metaphysical statements whose status would be threatened if the doctrine were actually correct. Nevertheless, precise formulations aside, the general idea of verification has had enormous influence on subsequent philosophers. According to some, such as Carl Hempel and Quine, what was wrong was the focus on *individual* statements. Instead, they suggested, influenced by Pierre Duhem (1906), it is whole systems of statements that scientists seek to verify in their empirical activities. What has come to be called the Duhem-Quine thesis, that a scientific hypothesis cannot be tested in isolation, indicates the shift that there has since been from atomistic to holistic conceptions of verification. In the case of Quine, this was famously accompanied by rejection of the analytic/synthetic distinction that lay at the basis of logical positivism (Quine, 1951). But this rejection was only intended to purge empiricism of untenable doctrines, not to repudiate empiricism altogether. For Quine, philosophy and science are continuous, and even analytic statements are subject to revision in the light of empirical research.

Other philosophers who were also broadly sympathetic to verificationism took different approaches. In Dummett’s work, for example, in which holism is rejected (cf. 1991: ch. 10), the concern has been to develop a systematically articulated theory of meaning based on the notion of assertion-conditions rather than on the classical notion of truth-conditions. A verificationist rather than traditional conception of truth lies at the heart of Dummett’s project, and he has

seen in mathematical intuitionism the model for a semantics based on verificationist truth. Dummett has been led to reformulate many traditional debates, such as those concerning our knowledge of the past and future, in terms of the opposition between realism and antirealism, a reformulation clearly influenced by the verificationism of the logical positivists. (See, for example, Dummett, 1993, which starts by paying homage to Ayer. Cf. also Wright, 1993: chs. 9–10.) Despite the flaws in its original articulation, then, verificationism remains an active – though controversial – force in contemporary philosophy.

See also: Analytic/Synthetic, Necessary/Contingent, and a Priori/a Posteriori; Distinction; Empiricism; Holism, Semantic and Epistemic; Limits of Language; Meaning: Overview of Philosophical Theories; Realism and Antirealism; Truth: Theories of in Philosophy.

Bibliography

- Austin J L (1962). *Sense and sensibilia*. Oxford: Oxford University Press.
- Ayer A J (1936). *Language, truth and logic*. London: Gollancz. [2nd edn., 1946.]
- Ayer A J (1936/1937). ‘Verification and experience.’ *Proceedings of the Aristotelian Society* 37. [Reprinted in Ayer (ed.) 1959, 228–243.]
- Ayer A J (ed.) (1959). *Logical positivism*. Glencoe, Illinois: The Free Press.
- Beane M (2004). ‘Carnap’s conception of explication: from Frege to Husserl?’ In Awodey S & Klein C (eds.) *Carnap brought home: the view from Jena*. Chicago: Open Court. 117–150.
- Berlin I (1938/1939). ‘Verification.’ *Proceedings of the Aristotelian Society* 39, 225–248. Reprinted in Parkinson G H R (ed.). *The theory of meaning*. Oxford University Press: Oxford. 1968. 15–34.
- Carnap R (1928). *Der logische Aufbau der Welt*. Berlin-Schlachtensee: Weltkreis-Verlag. [Translated in *The logical structure of the world*. George R A (trans.). Berkeley: University of California Press, 1967.]
- Carnap R (1932). ‘The elimination of metaphysics through logical analysis of language.’ *Erkenntnis* 2, 219–241. [Translated in Ayer (ed.), 1959, 60–81.]
- Carnap R (1932/1933). ‘Psychology in physical language,’ *Erkenntnis* 3, [Translated in Ayer (ed.), 1959, 165–198.]
- Church A (1949). ‘Review of *Language, truth and logic*, second edition.’ *Journal of Symbolic Logic* 14, 52–53.
- Coffa J A (1991). *The semantic tradition from Kant to Carnap*. Cambridge: Cambridge University Press.
- Duhem P (1906). *La théorie physique, son objet et sa structure*. Paris: Chevalier et Rivière. [Reprinted in Wiener P P (trans) (ed.) *The aim and structure of physical theory*. Princeton: Princeton University Press, 1992.]
- Dummett M (1991). *The logical basis of metaphysics*. Cambridge, MA: Harvard University Press.

-
- Dummett M (1993). 'Realism and anti-realism.' In *The seas of language*. Oxford: Oxford University Press.
- Friedman M (1999). *Reconsidering logical positivism*. Cambridge: Cambridge University Press.
- Friedman M (2000). *A parting of the ways: Carnap, Cassirer, and Heidegger*. Chicago: Open Court.
- Hanfling O (ed.) (1981). *Essential readings in logical positivism*. Oxford: Blackwell.
- Hempel C G (1950). 'The empiricist criterion of meaning.' *Revue internationale de philosophie* 4, 41–63. [Reprinted in Ayer (ed.), 1959, 108–129.]
- Hume D (1748). *An enquiry concerning human understanding*. Selby-Bigge L A (ed.) [3rd rd. rev. by Nidditch P H. Oxford: Oxford University Press, 1975.]
- Neurath O (1932/1933). 'Protocol sentences.' *Erkenntnis* 3. [Translated in Ayer (ed.), 1959, 199–208.]
- Quine W V O (1951). 'Two Dogmas of Empiricism.' *Philosophical Review* 60, 20–43. [Reprinted in a revised form in *From a logical point of view*. Cambridge, MA: Harvard University Press, 1st ed. 1953, 2nd ed. 1961, 20–46.]
- Richardson A (1998). *Carnap's construction of the world: the Aufbau and the emergence of logical empiricism*. Cambridge University Press: Cambridge.
- Russell B (1914). *Our knowledge of the external world*. Open Court: Chicago.
- Schlick M (1934). 'The foundation of knowledge.' *Erkenntnis* 4. [Translated in Ayer, 1959, 209–227.]
- Soames S (2003). *Philosophical analysis in the twentieth century* (2 vols). Princeton: Princeton University Press.
- Swartz R J (ed.) (1965). *Perceiving, sensing, and knowing*. New York: Doubleday.
- Uebel T E (1992). *Overcoming logical positivism from within*. Amsterdam: Rodopi.
- Urmson J O (1956). *Philosophical analysis: its development between the two World Wars*. Oxford: Oxford University Press.
- Wisdom J (1931–1933). 'Logical constructions.' Part I, *Mind* 40, Part II, 40; Part III, 41, 441–464; Part IV, 42, 43–66; Part V, 42, 186–202.
- Wright C (1986). *Realism, Meaning and Truth*. Oxford: Blackwell.
-

This page intentionally left blank

SUBJECT INDEX

NOTE

The index is arranged in set-out style with a maximum of three levels of heading. Major discussion of a subject is indicated by bold page numbers. Page numbers suffixed by T and F refer to Tables and Figures respectively. *vs.* indicates a comparison.

Cross-reference terms in italics are general cross-references, or refer to subentry terms within the main entry (the main entry is not repeated to save space). Readers are also advised to refer to the end of each article for additional cross-references - not all of these cross-references have been included in the index cross-references.

This index is in letter-by-letter order, whereby hyphens and spaces within index headings are ignored in the alphabetization. Prefixes and terms in parentheses are excluded from the initial alphabetization.

To assist the user, specific languages have main entries in **bold**. To save space in the index, subjects specific to a language (e.g. morphology, syntax) are found as subentries under the language and not the subject. Cross references to the specific subjects (e.g. syntax, phonetics) are assumed and not explicit.

The following abbreviation have been used:

L2 – second language

A

AAPC *see* Aristotelian-Abelardian predicate calculus

Aboutness, 334–335

Absolute tenses *see* Tense

Abstracta, 228, 545–546, 798, 799

Abstraction/abstract thought, 491, 581

nominalism, 521–522

workers in, Locke, 300

see also Case; Data and evidence; Object(s)

Accessibility hierarchy, 136

Acquaintance principle (Russell), 536, 639, 641, 643, 647–648

Acquisition of language *see* Language acquisition

Action

explanation, 224

intentionality, 3

Actionality *see* Aktionsart

Action sentences, 3–6, 34

Adams, R M, 580, 581–582

Adaptation

syntax, 207

Adequacy *see* Levels of adequacy/observation

Adjectives

autological/heterological, 557–558

gradable comparatives, 424

as indexicals, 324, 325

as natural kind terms, 501

predication *see* Predication, nonverbal

Adverb(s), 324

event arguments, 193

Advertising, 335

Agency

intentionality, 3

Agreement

verbs, 600

Aitchison, J, 150

Ajdukiewicz, Kazimierz, 681

Aktionsart, 194

Albert of Saxony, 522, 715–716

Algebra

Boolean, 36–39

Ambiguity

compositionality, 74

de dicto/de re distinction, 126–127

definite descriptions, 645

lexical *see* Lexical ambiguity

pragmatic, 512–513

of scope, 168, 193, 411–412, 493

quantifiers, 627

semantic, 358, 487, 574, 577–580, 646, 647, 689–690, 700

disjunction, 248

underspecification, 128

American linguistics, history, 785

American Sign Language (ASL)

apes, 198–199

Amis, K, 149

Analyticity

analytical hypotheses, 322, 631

linguistic, 12

Analytic philosophy, 6–9

language and social construction, 702

Analytic/synthetic distinction, 404, 405

logical positivists, 2, 16–17

workers in

Grice, 16

Kant, 1, 2, 13, 96–97

Kripke, 3

Quine, 2, 8–9, 14, 78–79, 97, 182, 812

Strawson, 16

Anaphora, 107–109, 575, 677–678

binding constraints, 602

bound-variable, 107, 160

bridging cross-reference, 108

context, 247

contradictions, 144

definition, 107, 144

deixis *vs.*, 144

dynamic predicate logic, 174

E-type (donkey), 108, 164, 169–171, 279–281

identity, 107

of laziness, 108

'linguistic' approach, 144

N-bar, 108

philosophical aspects, 19–21

Context-Dependent Quantifier

Approach, 20

definition, 19

Descriptive Approaches, 20, 156

Discourse Representation Theory, 19, 20

donkey, 20

Dynamic Logic Accounts, 21

referring expression, 19, 326

pragmatics, 144–146, 352

presuppositions, 592

process of recovery, 144

pronouns, 145, 763

VP-ellipsis, 108

Anaphoric reference, 19–21

Animal(s)

communication, 64

intentionality, 704

thought, 748

Animal communication

intention-reading skills, 198

Anscombe, G E M, 306, 307

Anti-functionalism, 561

Anti-individualism, E-language/I-language distinction, 179

Antirealism, 634–637

global, 636–637

Antony, L, association with Hornstein, N, 733

Antonyms/antonymy, 359

see also Hyperonymy; Hyponymy; Negation; Synonym(s)

Apes, gesture communication

sign language, 198–199

Apollonius Dyscolus, 497

a priori/a posteriori distinction

analytic/synthetic distinction, 3

necessary/contingent distinction, 46

see also Knowledge

Aquinas, Thomas, 415, 521–522

Äqvist, L, 685

Architecture of language system, 693, 694

Ariel, Mira, 136

Aristotelian-Abelardian predicate

calculus, 591

Aristotle, 25–27

biography, 25

fields of work

empiricism, 183, 184

figurative language, semiotics of, 231

meanings as mental states, 53

names, 28

origin of language, 552

reference, 636, 637

rhetoric, 780

sylogistic logic, 398

terms for word classes, 27

truth as correspondence, 777

universals, 520–521

writings

Categories, 26

De Interpretatione, 27, 585

Metaphysics, 26

Prior analytics, 26

Armstrong, D, 189, 334

Artificial intelligence (AI)

frame problem, 265

nonmonotonic inference, 524

planning theory, speech acts *see* Speech acts

Artificial languages, 97, 399

Asher, N, 616

Aspectual character *see* Aktionsart

Aspectual class *see* Aktionsart

Assertion, 28–30, 90–91

convention, 414

conversational challenge, 29

denials, 512–513

direction of fit, 28

fiction, 227

intention, 334, 532

liability to error, 28–29

lying, 415

- Assertion (*continued*)
 moral norms, 413–415
 No Proposition View, 186–187
 presuppositions, 29
 private language, 606
 proto-assertions, 335
 sincerity, 30–30, 79, 187, 188, 617
 warranted assertibility, 635
- Associationism and connectionism, 183
- A-theory, 741, 742
- Atomism
 compositionality, 71–72
 conceptual, 329
 logical
 Russell, 8, 402–403
 Wittgenstein, 8, 405
 semantic, 296, 297, 420
 workers in
 Russell, 8, 402–403
 Wittgenstein, 8, 405
- Attention
 in infants
 joint *see* Joint attention
 nonlinguistic goals, 429
 joint *see* Joint attention
- Attitude, propositional, 617
 ascriptions, 613–617
 attributions, 610
 as context sensitive, 613–614, 614–615
de dicto/de re distinction, 614–615
 definite descriptions, 153
 generality constraint, 734, 735
 opacity problem, 613–614
 relational analysis, 613
 representation, 656
 sentential structure, 442
 tacit knowledge, 733, 735
see also Beliefs
- Attrition, L2 *see* Second language attrition
- Auditory gestures *see* Apes, gesture communication
- Austin, John Langshaw, 28, 437, 801–802
 fields of work
 ordinary language philosophy, 49, 548, 551
 pragmatics, 312–313, 792–793
 speech-act theory, 478, 705
see also Speech acts, Austen publications
How to do things with words, 478
- Autonomy, 258, 263
 formalism, 258
 linguistic form, 378
 mentalist view, 258
 semantics, 350
see also Associationism and connectionism; Binding; Constructivism; E-language *vs.* I-language; Functionalism; Language of thought; Modularity; Nativism; Rationalism; Scope
- Auxiliary copying, definition, 388
- Auxiliary verbs
 English, 753–754
- Availability Principle, 690
- Ayer, A J, fields of work
 analyticity, 17
a priori knowledge, 14, 15, 96–97
 definition, 10–11, 16
 deflationism, 142
 expressivism, 635
 necessary truth, 10–11, 11–12, 15
 sense datum theory, 549
 verificationism, 810–811
- B**
- Bach, K, 527, 529, 616, 691
 Harnish, R M, association with, 334
- Background, non intentional, 216
- Bacon, Francis, Lord Verulam, 183
- Baker, Mark, 332
- Bakhtin, Mikhail Mikhailovich
 fields of work
 pragmatics, 317
- Barber, A, 647–648
- Bare noun phrases, 135
 numerals, 626
- Bare plurals
 generic reference, 290, 292
- Barker, S J, 335
- Barnes, J, 744–745
- Barsalou, L, 200
- Barthes, Roland
 fields of work
 Saussurian theory of the sign, 670–671
- Barton, G E Jr, 263
- Barwise, Jon, 700
 co-workers and associated workers
 Cooper, R, 681–682
 Perry, J, 461, 463–464
- Bealer, G, 16–17
- Becoming, temporal, 741, 742
- Bedeutungsfeld *see* Semantic field
- Behabitives, speech acts, 709
- Behaviorism, 31–35, 346, 374, 561
 Bloomfield, Leonard, 476
 conversation analysis, 376
 determinacy of meaning, 322, 323, 607–608, 771
 epistemology, 183
 history of, 476
 methodological, 31, 631–632, 749
 operational, 31–32
 ontological, 749
 proper names, 54–55
 psychological
 emphasis on learning, 35
 radical, 32
 rejection, 52–53, 180
 thought and language, 222, 613
 varieties, 52
see also Quine, Willard van Ormand
- Belief and Sincere Assertive Utterance Problem, 186–187, 188
- Belief-Revisability Argument, 14–15
- Beliefs
 assertions, 30–30
 communicating, 64, 66–67, 773
 confirmation holism, 14
de dicto/de re, 493–494
 ethnomethodological concerns
 externalism, 190
 indexical, 191
 make-believe, 227, 228, 229
 propositional attitudes, 166
 sentential structure, 442
 testimony, 745, 745*t*
 truth-value, 774
- Belnap, Nuel, 4
- Benceraff, P, 619
- Benveniste, Émile, 784
 fields of work
 Saussurian theory of the sign, 669
- Berkeley, George, 299, 300, 548
- Berry's Paradox, 367, 558
- Über Sinn und Bedeutung, 220
- 'Bestrangement'
 pragmatics, 319
- Bickerton, Derek, 205, 207, 208
- Bierwisch, M, 355
- Bilingualism, 120
see also Interlanguage
- Binding
 constraints, 600
 anaphora, 602
 Italian, 601
 pronouns, 603
- Binding scope, 279–281
 ambiguity *see* Ambiguity
 definition, 677
 dependency, 677
 donkey anaphora, 170
- independence, 677
 semantic aspects, 677–679
- Binding theory, 391–392, 567, 621
 Principles and Parameters Model, 602
 as science, 393
see also Anaphora; Coreference; Deixis; Exophora; Pronouns; Reflexives
- Bivalence principle, 635
- Bivalent logic, 513
- Black, Max, 448
 figurative language, semiotics of, 234
- Blackburn, S, 299–300, 334, 608
 Simmons, K, association with, 142
- Blind application of transformations, 765
- Block, N, 296
- Bloom, P, 205, 209
- Bloomfield, Leonard, 374, 376, 566, 786
 fields of work
 behaviorism, 476
- Boas, Franz, 785
 fields of work
 anthropology, 475
 publications, *Handbook of American Indian Languages*, 475
- Boghossian, P, 13, 14–15, 16–17
- Bonjour, L, 16–17
- Boole, George, 36–39
 fields of work
 logic, 399
- Boolean compounds, 625
- Bopp, Franz
 fields of work
 comparative philology, 473
- Boroditsky, L, 200
- Borrowing, 577
- Botterell, A, 647–648
- Bound-variable anaphora, 107, 160
- Brandom, R B, 29, 658
- Bresnan, Joan, 262
- Bridging cross-reference anaphora, 108
- Broad, C D, 741
- Brøndal, Rasmus Viggo, 784
- Brown, Gillian, 527
- B-theory, 741, 743
 date-analysis account, 742
 token-reflexive account, 742
- Burge, Tyler, fields of work
 demonstratives, 338
 externalism, 179, 535
 naturalism, 508
 proper names, 54
- Buridan, John, 522, 715–716, 717
- Burleigh, Walter, 169–170
- Burling, R, 208
- Burton-Roberts, N, 563
- Byrne, R, 198
- C**
- CA (conversation analysis) *see* Conversation analysis
- Calculus
 logical, 398–399
see also Lgr-calculus; Propositional calculus
- Lgr-calculus
 linguistic applications, 491
- Cancellation conditions, 590, 591–592
- Cancellation of defaults, 129
- Candidate competition, Optimality Theory, in morphology *see* Optimality Theory
- 'Career of metaphor' theory, 453
- Carnap, Rudolf
 fields of work
a priori knowledge, 14, 15, 96–97
 attitude ascriptions, 615
 definition, 16
 epistemology, 507, 810
 indirect reference, 697
 logic, 8–9, 12, 405, 717
 public language, 750–751
 state descriptions, 682
- Carr, P, 562–563

- Carston, R
 fields of work
 conversational implicature, 529, 689
 explicature, 690, 692
 Linguistic Direction Principle, 691
 mentalism, 693
 pragmatic determination, 690
 pragmatic intrusion, 693, 694
 relevance theory, 439
- Cartesianism
 the mind, 632
 private language argument, 606, 607
see also Descartes, René
- Cartwright, R, 142–143, 804
- Carus, Titus Lucretius, origin of language, 552
- Case
 relations
 Icelandic, 599–600
- Castañeda, H N, 191, 192
- Categorical grammar, 22, 24, 680, 681
- Categories, 26
 child development of
 lexical meaning and, 428, 431
 concepts, 78, 79
 conceptual atomism, 80
 generic, 349
 metaphor, 453
 theory-theory, 79–80
- Categories* (Aristotle), 26
- Category mistakes (logic), 8, 34, 513
- Causal-historical theory of reference, 88
 descriptivism, 41
 direct reference, 158, 159, 641, 643
 proper names, 41, 42, 159, 641
- Causal theories of reference, 41–44, 610, 643
 descriptivism, 41
 externalism, 42
 information-theoretical approach, 642
 Kripke, 42, 503, 610, 641–642
 natural kind terms, 501
see also Causal-historical theory
- Causal theory of meaning, 43
- c-command, 392–393, 412, 678
 definition, 390–391
see also Scope
- CDQ *see* Context-Dependent Quantifier Approach
- Central processing system, 562
- Chakrabarti, A, 744, 745
- Chalmers, D, 606, 796–797
- Chan (Zen), 365–366, 367
- Character/content distinction, 44–46
 demonstratives, 44, 159, 216, 795
 indexicals, 44, 159, 216, 795
 philosophy, 660
 semantics, 157, 324–325, 773
- Characteristic functions, verbs, 249–251, 252
- Charity principle, 67, 322
 radical interpretation, 631, 632, 633–634, 772–773
- Chierchia, Gennaro, 21, 155, 687
 predication, 587
- Child-directed speech (CDS)
 lexical meaning development, 427–428
- Children
 coining words, 113
 intention, 197
see also Infant(s); Language acquisition
- Chinese, 411
- Chisholm, R, 191
- Chomsky, Noam, 339, 503
 cognitivism, 51
 fields of work
 adequacy, 375
 analyticity, 16
 behaviorism, 32
 conceptual-intentional system, 693
 E-language (externalized), 179–182
 epistemology, 733, 736–737
 evolution, 205
 idealization, 375, 379–380
 innateness, 331, 332
 key principles, 374
 language acquisition, 185
 linguistic convention, 179
 linguistics, 566, 567
 logical form, 411, 412, 693, 694
 naturalism, 508
 noncommunicative uses of language, 268
 Principles and Parameters Model, 24
 rationalism, 379–380
 reference theory, 53, 642
 rejection of behaviorism, 52–53
 semantic indeterminacy, 323
 standard language, 527–528
 type/token distinction, 799
 generativism, 477, 787
 generative grammar, 347
 generative phonology
 grammar
 transformational, 753
 publications
Syntactic structures, 753–754, 787–788
 rejection of behaviorism, 52
 syntax, 23, 24–25, 206, 378–379, 382, 694
see also Formalism/formalist linguistics;
 Generative grammar; Philosophy of linguistics
- Christophersen, Paul, 132
- Church, Alonzo, 461, 697
- Circumscription, 525
- Circumstance of evaluation, 44, 795
- Clark, Andy, 751
- Clark, Herbert H, 529
- Clash
 cooperative principle, 101
- Class
 Boolean logic, 399
 type, 370
 words *see* Word classes
- Class-inclusion model, metaphor, 453
- Clause(s)
 types, 496–500
 embedded *see* Embedded clauses
 identification, 498
 illocutionary force, 497, 500
 illocutionary point, 497–498, 499–500
 primary illocution, 497
 relative *see* Relative clauses
 subsentences, 498, 499–500
see also Speech acts
- Cleft sentences, 247
- Closed world assumption, 524–525
- Closure, semantic, 558
- Clothing
 bodies, 58
 'dress codes', 59
 fashion, 62
 history, 61
 nudity, 62
- Co-composition, 693
- Coercion
 aspectual, 693
- Cognitive dependency, 424–426
see also Lexical conditions; Polysemy
- Cognitive dominance, 200, 200f
- Cognitive frames *see* Frame(s)
- Cognitive grammar, 185
- Cognitive linguistics, 51–52
- Cognitive mechanisms, metaphor, 451, 452–453
- Cognitive principle of relevance, relevance theory, 650
- Cognitive representation, lexicon *see* Lexicon
- Cognitive science, 375
 language processing *see* Language processing
 philosophy of language, 47–56
 behaviorism, 33–34, 35
 ideal language philosophy, 47
 ideas, 184
 indexicality, 326–327
 innateness, 327–328
 language processing, 23
 new philosophy of language, 54–55
 ordinary language tradition, 47
see also Artificial intelligence; Cognitivism
see also Behaviorism; Cognitivism
- Cognitive semantics
 cognitive defaults, 128–131
 cognitive dependency, 424–426
 dependency *see* Cognitive dependency
 polysemy, 579
- Cognitivism
 computational and representational theories of mind, 53, 443, 605–606, 613
 emotion, 229
 I-language, 180
 objectivity in moral discourse, 539, 540, 542
 three-level research program, 52, 52t, 53
see also Chomsky, Noam; Connectionism;
 Fodor, J; Minsky, Marvin
- Cohen, L J, 689–690, 691
- Cointersective determiners, 626
- Co-intersective determiners *see* Universal quantifier
- Color, 199
see also Cognitive semantics; Prototype semantics; Universals
- Commissives
 speech acts, 709
- Common ground, 85–87
 definition, 85
see also Presupposition
- Common knowledge, 85
- Communal common ground, 85
 linguistic aspects, 86
- Communicated acts, relevance theory, 652
- Communicated Proposition View, 188
- Communication
 competence *see* Pragmatics
 coordination, 97, 98
 intention, 64–65, 97, 98–99, 215, 216, 334, 693
 linguistic, 66, 299
 nonlinguistic, 64–65
 philosophy, 64–68, 323, 607
 private thoughts, 66, 301
 semiotics, 57–64
 code-model approaches, 62
 functionalist reading, 59
 the model of communication within the functionalist reading, 61f
 model of textual functions, 61f
 Pierce, Charles Sanders, 62, 62f
 signification, 57, 58
 translation *see* Translation
see also Discourse
- Communicative intention, relevance theory, 649–650
- Communicative principle of relevance, 650
- Comparative grammar, 596
see also Comparative philology
- Comparative philology
 definition, 472–473
 history of, 473
 origins, 473
 workers in
 Bopp, Franz, 473
 Grimm, Jacob Ludwig Carl, 473
 Jones, William, 473
 Müller, Friedrich Max, 473
 Rask, Rasmus Kristian, 473
see also Comparative grammar
- Comparatives, 68–71, 625–626
 complex, 70
 definition, 68
 similarity, 83
 standards, 69
see also Ambiguity; Quantifiers; Vagueness
- Comparison models, metaphor, 453
- Compensatory lengthening *see* Quantity
- Competence, 375, 561, 567, 568, 688
 acquisition, 331
 compositionality, 730
 first-person authority, 632–633
 innateness, 332
 performance, 24–25, 688, 693
 pragmatics, 24, 688
 semantic, 246
- Complementizer phrase, definition, 390–391
- Complex events, 194–195
- Complex expressions, 75–78, 81–84, 246
- Complex sentences *see* Sentence(s)

818 **Subject Index**

- Componential analysis, 357–358
 Composition, enriched, 690, 693, 694
 Compositionality, 71–75, 75–78, 220, 279–281, 354, 688, 728
 concepts, 79, 80
 context principle, 90, 92, 613–614
 definition, 75
 denoting concepts, 152, 153
 epistemic, 730
 formal language, 22, 23–24
 intensionality merger, 128, 129, 130
 magic of language, 73, 771
 as mathematical function, 72–73, 74
 metaphysical, 71–72
 Montague semantics, 487
 ontic, 730
 reverse, 730
 semantic, 73, 296, 613–614, 615–616, 773
 semantic value, 680
 speaker-meaning, 334
 systematicity, 728, 730
 truth conditions, 772
 see also Noncompositionality
 Computational semantics, 773
 CON *see* Optimality Theory
 Concepts, 78–81
 classical theory, 80
 complex, 80
 conceptual atomism, 329
 denoting, 152, 153
 direct reference theories, 606
 empiricism, 184
 generality constraint, 734, 735
 incompleteness, 691
 innate *see* Nativism; Concepts
 language, 751
 learning, 78, 79, 328–329
 lexical, 328
 lexical concepts, 196–197, 200, 200f, 201
 medieval nominalism, 523
 nativism, 80, 327, 329–330, 443
 natural kind, 329, 523
 probabilistic theories, 80
 reference, 79, 638
 Conceptual encoding, relevance theory, 651
 Conceptual formation rules, 352
 Conceptual integration, 201, 201f
 Conceptualism, universals, 520, 799
 Conceptual meanings, procedural *vs.*, 438–442
 Conceptual metaphor theory, 202
 Lakoff, G & Johnson, M, 235
 psychological studies, 454
 examples, 455
 Conceptual structure, 349–356
 definition, 349
 primitives, 353, 355
 Condillac, Etienne Bonnot de, 553f
 fields of work
 origin of language, 553
 Conditionals, 81–84, 256
 definition, 81
 see also Counterfactuals; Inference; Modal logic; Possible worlds
 Conditioning
 classical, 32, 35
 operant, 32
 Confirmation holism, 14, 34, 183–184, 295
 Conjunction, 254–255
 nouns, 576
 Conjunction Analysis (for presupposition), 592
 Connectionism
 knowledge acquisition, 35, 183
 language processing, 53, 443, 444
 Connotation, 221
 denotation, 609
 denotation *vs.* *see* Denotation
 negation, 517
 proper names, 158
 Consequence, logical *see* Logical consequence
 Conservativity (logic), 624
 Consistency
 negation, 511
 Constatives
 speech acts, 705
 Constituency *see* Phrase structure (constituency)
 Constituents/constituency *see* Phrase structure (constituency)
 Constituent structure-functional structure relation
 see C-structure
 Constraint(s)
 models *see* Constraint-satisfaction models
 syntactic
 transformations, 598
 see also Island constraints
 Constraint-satisfaction models
 metaphor, 451–452
 Constructivism, 543, 636–637
 emergent modularity, 562
 Contact-induced change, types *see* Types
 Content
 associated, 609–610, 610–611, 612
 causal theory, 224, 329
 conceptual, 78
 enriched/loosened, 690
 externalist view, 222–225
 intentional, 536
 locutionary *vs.* illocutionary, 689
 narrow *vs.* wide, 46
 naturalistic account, 224, 508
 propositions as, 617, 620
 teleofunctional account, 224
 see also Explicatures; Implicature; Implicatures
 Context, 44, 46
 context set, 795–796
 definition, 90, 138, 140
 indexicals and demonstratives, 44, 191–192, 582–583, 691
 intensional *vs.* extensional, 125
 meaning, 73, 74, 324, 405, 528, 688
 metaphor, 447
 opaque, 125, 126, 435, 613–614
 pragmatics, 310
 sensitivity
 interrogatives, 687
 pragmatic determinants, 582–583
 propositional attitudes, 613–614, 614–615
 truth-value, 682–683, 775, 794
 transparent, 125, 614
 vagueness, 808–809
 Context-bound knowledge, 425
 see also Common ground
 Context-Dependent Quantifier Approach, 20
 Context-free Grammar (CFG), 263
 Context principle (CP), 76, 88–94
 compositionality, 90, 92, 613–614
 interpretational *vs.* psychological reading, 89, 93
 metasemantic reading, 88, 91–92
 methodological reading, 88, 91
 semantic holism, 297
 Contextual(s), 325
 relevance theory, 650–651
 Contextualization
 epistemology, 94–96, 189
 Contradictions, anaphora, 144
 Contraries, negation, 511
 Contrast
 conversational, 427
 Control
 definition, 766
 Conventionality
 conversational, 427
 indirect speech acts, 710
 Conventions, 96–100
 moral norms, 414
 non-standard use, 527
 private, 606
 of truthfulness and trust, 98
 workers in
 Carnap, 404–405
 Chomsky, 179
 Grice, 99, 773–774
 Lewis, 180
 Convergence
 objectivity, 539
 Conversation
 analysis *see* Conversation analysis
 assertions, 29
 child development
 lexical meaning, 427
 common ground, 429
 context, 795–796
 contrast, 427
 conventionality, 427
 Cooperation Principle, 688–689
 cooperativity, 427, 429
 Grice's maxims, 427
 grounding *see* Grounding
 manner, 427
 maxims, 447, 532, 647–648, 688–689
 norms, 29–30
 quality, 427
 quantity, 427
 relation/relevance, 427
 see also Discourse; Implicature; Pragmatics
 Conversational implicature *see* Implicature
 Conversational postulates, indirect speech acts, 711
 Conversation analysis, 376
 Cooper, R, association with Barwise, J, 681–682
 Cooperative principle, 100–106, 318, 421
 conversation, 427, 429
 criticisms, 102
 maxim problems, 103
 terminology, 102
 definition, 100
 influences, 103
 gender studies, 105
 grammar, 104
 neo-Gricean pragmatics, 104
 pedagogy, 105
 politeness theory, 104
 question processing, 105
 maxim failure, 101
 clash, 101
 flouting/exploiting, 101
 opting out, 101
 violation, 101
 maxims *see* Maxim(s)
 maxims of cooperative discourse, 100
 Coordination, 254
 quantifiers, 484
 transformation, 756
 see also Ellipsis; HPSG; Island constraints; Phrase structure (constituency); X-bar theory
 Coordination problems (physical), 97, 98
 Copenhagen School
 history, 784
 Coping strategies
 lexical acquisition, 431
 Corballis, M C, 205
 Coreference, 107–109, 220
 definition, 107
 see also Anaphora
 Corpus linguistics
 confounds, 118
 corpora as data source, 117
 corpus data: definition, 117
 evidence, 117
 limitations, 117
 non-occurrences, 117
 presence of forms, 117
 quantitative data, 117
 written language, 118
 Correspondence theory of truth, 142, 540, 775, 777–778
 Cotenability, 110
 Counterfactuals, 81, 109–112, 470–471
 definition, 109
 see also Conditionals; Possible worlds, semantics
 Counting
 measurement, 419
 'Counting' quantifiers, 677, 678–679
 Count noun phrases, 36, 37
 Count terms, generic reference, 290
 Cours de linguistique générale (de Saussure), 665
 Course in general linguistics (CLG)
 (de Saussure), 474
 Cowart, W, 121–122
 CP *see* Context Principle

- Craig, W L, 741
 Crain, S, co-workers and associated workers
 Gualmini, A, 332
 Pietroski, P, 332
 Cratylus *see* Plato
 Creativity, 112–115
 academic terminology, 113
 behaviorism, 32, 33
 coining words, 112
 children's coining, 113
 word formation, 112–113
 language change, 112
 language revival, 113
 Irish Gaelic, 114
 Modern Hebrew, 113–114
 modern vocabulary, 113–114
 writing systems, 114
 lexicography, 112
 linguistic, 73, 770
 meaning extension, 114
 metaphor, 446, 448
 speakers' needs, 112
 syntactic units and combinations, 114
see also Compositionality; Semantic change;
 Word formation
 Creolization, 120
 Creswell, M J, 338
 Crimmins, M, 615
 Croft, W A, 268–269, 270
 Crosslinguistic studies/variation
 anaphora
 binding constraints, 602–603
 comparatives, 70
 definite/indefinite noun phrases, 136
 lexicalization, 426–427
 predication, 587–588
 Crosthwaite, J, 448
 CS *see* Conceptual structure
 C-structure
 transformational grammar, 757–758
see also Configurationality
see also X-bar theory
 Currie, G, 227, 228
 Ccount nouns *see* Noun(s)
- D**
- Data and evidence, 117–123
 data, evidence, and theory, 122
 derivational theory of complexity (DTC), 122
 ethnographic purposes, 119
 kinds of, 120
 bilingualism, 120
 creolization, 120
 historical change, 121
 L2, 120
 language acquisition, 120–121
 language disorders, 121
 performance errors, 121
 types, 121
 observer's paradox, 119
 rationalism *vs.* empiricism, 379
 sociolinguistic purposes, 119
 sources of, 117
 brain measurements, 120
 corpora, 117
 experiments, 119
 fieldwork, 119
 grammaticality judgments, 118, 119
 variability, 121
 confounds, 121–122
 I-languages, 121
 patterns of responses, 121–122
see also Bilingualism; E-language *vs.* I-language;
 Language acquisition; Psycholinguistics;
 Sociolinguistics; Transformations
 Davidson, Donald, 66, 697
 fields of work
 action sentences, 4–5
 belief and meaning, 773
 convention, 99, 414
 epistemology, 190
 event arguments, 192–193, 193–194, 740
 identity criteria, 303
 indirect speech reports, 615
 interpretationalism, 323–324
 interrogatives, 685
 mental representation, 657–658
 radical interpretation, 631–632, 772–773
 same-saying, 339
 semantic innocence, 336–337, 685
 slingshot argument, 461, 462
 theory of truth, 76
 truth conditions, 772, 775
 use theories of meaning, 8–9, 296,
 437, 528–529
 Davies, M, 20
 Humberstone, L, association with, 796
 Declarative clause, 496–497, 498, 498*t*, 499
 Declarative mood, 497
 Declaratives
 speech acts, 709
 Decomposition *see* Lexical decomposition
 Decreasing generalized quantifiers, 624
 De dicto/de re *vs.*, 124–127, 493–494, 614–615
see also Propositional attitudes
 Default(s)
 cancellation, 129
 possessives, 130
 semantics, 128–131
 definition, 128
see also Compositionality; Presupposition
 Definite(s), 131–138
 definition, 131
 noun phrases, generic reference, 290
 quantified noun phrases, 624
see also Pronouns
 Definite descriptions *see* Descriptions, definite
 Definite determiners, 625
 Definiteness effect, 134
 Definition, 138–141
 abbreviative, 138
 analyticity, 10–11, 16, 78–79
 Berry's Paradox, 558
 descriptive, 138
 enumerative, 140
 explicative, 138
 explicit, 139
 explicit stipulative, 17, 405, 568, 610
 implicit/contextual, 17, 90, 138, 139, 140, 152
 lexical, 138
 necessary truth, 12, 15
 ostensive, 140
 probabilistic theory, 79
 recursive/inductive, 139
 Richard's Paradox, 558
 stipulative, 138, 139
 theoretical terms, 140–141
 Definition (lexicography) *see* Lexicography
 Deflationism, 141–143
 meaning, 143, 803
 trut, 142, 776–777
see also Minimalism
 Degrees of intentions, 128–129
 De Interpretatione *see* Aristotle
 Deixis
 anaphora *vs.*, 144
 definition, 144
 pragmatics, 144–146
 symbolic meanings, 145–146
 Democratization, Internet impact
see Internet
 Demonstrative(s)
 character/content distinction, 44, 159, 216, 795
 complex, 324
 context, 44, 582–583
 embedded, 337–338
 externalism *vs.* internalism, 223
 Kaplan's theory, 171
 object-dependent thought, 535–536, 538
 pure indexicals, 325
 reference, 45, 637, 648
 semantics, 338
 surrogates, 171
 Denial, 516
see also Negation
 Dennett, D, 51, 323–324, 509, 633
 Denotation, 220
 aboutness, 334–335
 connotation *vs.*, 609
 count/non-count nouns, 419
 facts, 457
 fiction, 228
 intention, 335
 natural kind terms, 501, 502,
 503, 504
 workers in
 Russell, 615
 Derivational Complexity Hypothesis, 757
 Descartes, René, 66, 553*f*, 561–562
 fields of work
 category-mistakes, 8
 knowledge skepticism, 635
 origin of language, 553
 psychological explanations, 32
 Description and prescription, 146–151
 authority
 language academies, 147
 of usage, 148–149, 149–150
 definition, 146–147
 grammar, 146–147, 149, 150
 corpus-based, 149, 150
 grammaticalization, 150
 history, 147
 innovation, 148
 language change, 148, 150
 language planning, 150
 lexicography, 147–148, 150
 modern linguistics, 149
 objectivity, 150
 punctuation, 149, 150
 responsibility, 150
 vocabulary, 147–148
see also Descriptivism; Grammar; Language as
 an object of study; Linguistics; Linguistics
 as a science
 Descriptions, definite, 151–157
 ambiguity thesis, 645, 647–648
 anaphora *as*, 20, 156
 disguised, 49, 54, 228–229, 646
 incomplete, 154, 647–648
 non-count nouns, 417
 object-independence, 536
 proper names, 610
 referential *vs.* attributive use, 644–648
 Donnellan, 644–645
 pragmatic approach (Kripke), 154, 215
 as referring expressions, 154, 215, 433,
 637, 647–648
 rigid designation, 659, 660
 Russell's theory, 125, 139, 417, 583
 disguised, 228–229
 incomplete descriptions, 154, 647–648
 Kripke, 646
 logical analysis, 8
 referential *vs.* quantificational, 464, 776
 sentences *as* truth-bearers, 776
 Strawson, 8, 154, 776, 801
 sense and reference, 696
 Descriptions, indefinite, 151, 152
 Description theories *see* Descriptivism
 Descriptive linguistics, 376
 Descriptive theory, speech *see* Speech
 Descriptivism
 causal-historical theory, 41
 competence, 502, 503
 epistemic problem, 640
 foundational, 502, 504
 meaning, 189, 611
 modal problem, 640
 natural kind terms, 501
 proper names, 638, 660
 property, 502, 503, 504
 reference, 643, 796–797
 semantic problem, 640
 workers in
 Frege, 41, 609–610, 612, 638, 640
 Quine, 189–190
 Russell, 49, 54, 609–610, 638
 Designation, rigid *see* Rigid designation
 Determinancy, 279–281

- Determinants, pragmatic *see* Pragmatic determinants
- Determination, compositionality, 72
- Determiners, 251, 625
see also Quantifiers
- Development of language *see* Language development
- Devitt, M, 155
- Dewey, John, 183–184
- Dialect(s)
definite/indefinite noun phrases, 133
language *vs.*
non-standard, 527–528
- Dialetheism, 511, 512, 559
- Dialogism
Saussurian linguistics, 668
- Dichotomies, philosophical *see* Analytic/synthetic distinction; *a priori/a posteriori* distinction; Necessary/contingent distinction
- Differential comparatives, 70
- Dik, S C, 270
- Direct access theories, irony, 341–342
- Direction of fit, 28
- Directives, speech acts, 709
- Directness, speech acts, 712–713
- Direct reference theory, 157–160, 606
causal-historical theory, 158, 159, 641, 643
definition, 615–616
indexicals and demonstratives, 45, 158, 325
Millian theories, 42, 143, 158, 610, 612, 771
proper names, 157, 158, 228, 433, 435–436, 612, 639
truth conditions, 158, 159, 434
- Direct speech acts, 710
- Discourse
direct, sense and reference, 435
domain, 590, 593
definite/indefinite noun phrases, 136
personal common ground, 87
see also Anaphora; Metaphor; Metonymy; Presupposition
fictional, 227–230
markers, 438–439
procedural analysis, 441
see also Conversation
- Discourse anaphora
referents, 173
- Discourse referents, 173, 176
- Discourse representation structures, 163
definition, 166
propositional attitudes, 166
- Discourse Representation Theory, 20, 128, 160–168, 171, 173, 616, 683
definition, 161
- Discursive psychology, media discourse *see* Media discourse
- Disjunction, 248, 254–255
- Displacement, 197
- Dispositionalism, 663
- Disquotationalism, 779
- Distinctions, philosophical, 10–19
- Distinctions of meaning, Plato, 570
- Distinctive feature analysis *see* Markerese
- Distinctness criteria, 418–419
- Distributionalism
Harris, Zellig S, 477
history of, 477
- Distributional restrictions *see* Optimality Theory
- Division of pragmatic labor *see* Implicature
- DL *see* Dynamic Logic Accounts
- Domain(s)
specificity
innateness, 329–330, 331
modularity, 331–332
- Domain-based knowledge *see* Encyclopedic knowledge
- Dominant language
see Lingua francas
- Donald, M, 197
- Donkey anaphora *see* E-type (donkey) anaphora
- Donnellan, K
ambiguity thesis, 154, 645, 647–648
causal theories, 41–42, 159, 640
expression *vs.* speaker meaning, 215, 528–529
sense and reference, 698
- Double bar notation, 249–251
- Downward entailment, 257, 515–516
- Downward monotonicity, 515–516
- Dowty, D, 681
- DPs *see* Determiners
- DR *see* Direct reference
- Dress codes, 59
- Dretske, F, 42, 507, 508, 509
- DRSs *see* Discourse representation structures
- DRT *see* Discourse Representation Theory
- Dthat, 171–172
definition, 171
- Dualism
substance, 613, 748
- Dual-stage processing, relevance theory, 651
- Duhem, P, 14, 34, 812
- Dummett, Michael Anthony Eardley
fields of work
context principle, 89
deflationism, 779
realism/antirealism, 634–635, 636
sense and reference, 698
sentences as truth-bearers, 90–91
verificationism, 810, 812
- Duponceau, Pierre Etienne
fields of work
linguistic anthropology, 475
- Dutch
Frames of Reference, 564–565
- Dynamic Logic Accounts of anaphora, 21
- Dynamic predicate logic, 76, 174, 470
existential quantifier, 174
- Dynamic semantics, 172–177
definition, 172
see also Anaphora; Discourse representation theory; Montague semantics
- E**
- Echo, 516, 518
relevance theory, 652
- Ecrits de linguistique générale (de Saussure), 474
- Egocentricity, indexicals, 191
- Einstein, Albert, 2
- E-language *vs.* I-language, 179–182, 339, 347, 561, 736
individualism *vs.* anti-individualism, 179
- Eliminativism, 613, 636–637
- Ellipsis, 23, 93, 583, 692
see also Coordination; Island constraints; Minimalism
- Embedded clauses
negation, 516–517
transformation, 755–756
- Embodiment, 199
- Emergentism, 274
- Emilian *see* Italian
- Emotion
fiction, 229
- Emotivism, 542
- Empiricism, 183–185, 379, 561–562, 563–564
~a priori knowledge, 12, 96–97
Belief-Revisability Argument, 14–15
concept, 184
data, 380
epistemology, 183, 184, 436, 502
innate ideas, 327, 329
language, 331
logical, 96–97, 295
meaning, 299
semantic, 184, 185
see also Naturalism
- Empty names, 185–188
examples, 48, 185, 610
fiction, 185, 228, 229
Fregeanism, 186
Gappy Proposition View, 186, 188
negative existentials, 212–213
new philosophy of language, 56
ordinary language, 49
workers in, Russell, 48
see also Proper names
- Encapsulation, 562
- Encyclopedic knowledge, 349–350, 355
cognitive dependency, 424–426
- Endangered languages, 119
- Endurance, diachronic identity, 303–304
- Engdahl, E, 687
- English
auxiliary verbs, 753–754
bare plurals, generic reference, 292
- Entailment relations, 482–485, 486
formal semantics, 246–247, 252–253
modal logic, 468
presuppositions *see* P-entailments
- Epicurian philosophy
origin of language, 552–553
- Epistemology, 189–191, 560
contextualism, 94–96, 189
descriptive theories, 189
empiricist theories, 183, 184, 436
externalism, 224
future tense, 276
holism, 190
intuitions, 183
logic, 401–402
naturalized, 11, 507
rationalism, 183, 184
skepticism, 94, 189, 190, 635
verificationism, 189, 295
see also a priori/a posteriori distinction; Knowledge
- Equi noun phrase deletion, 758–759, 764
- Equivalence
logical, 459, 460–461, 462–463
see also Identity
- Error(s)
assertions, 28–29
immunity to, 305–308
language acquisition, 331
moral, 540–541
non-standardized speech, 527
- Error theory, 542, 543–544, 636
- Essentialism
metaphysical, 523
psychological, 329, 501, 521
- E-type (donkey) anaphora, 108, 145, 164, 169–171, 279–281
definition, 169
dynamic predicate logic, 174
- Eubulides of Miletus, 557, 591
- Evaluation
circumstance, 44, 795
- Evans, Gareth
fields of work
anaphora, 19, 20
causal theories of reference, 43, 611, 641
contingency, 796
generality constraint, 734
I-utterances, 306–307
sense and reference, 698
singular terms, 158
vagueness, 807–808
- Evans, V, 202
- Event(s)
action sentences, 4
adverbs, 4
- Event-based semantics, 192–196
definition, 192–193
see also Discourse representation theory; Situation semantics
- Event nominals, 38
- Event structure, 361
- Event-token reification *see* Davidson, Donald
- Eventuality, hidden argument, 193
- Event variables, 193–194
- Evidentiality
tense interaction, 276
- Evolution of language *see* Origin and evolution of language
- Evolution of semantics *see* Semantic(s), evolution of
- Evolution of syntax *see* Syntax, evolution of
- Exaptation, 205
spandrels, 205, 207

- Excitives, speech acts, 709
 Exclamatives, 500
 Excluded middle, law of, 279–281
 Externalism *see* E-language *vs.* I-language
 Existence, 6–7, 211–214, 717
 internal/external questions, 405
 Meinong-Russell view, 636
 as property, 6–7, 212, 214
 see also Ontology
 Existential constructions, definiteness effect, 134
 Existential determiners, 625, 626
 Existential negatives, 6–7, 212, 213, 610, 636, 696–697
 Existential quantifier
 binding scope, 193
 dynamic predicate logic, 174
 Existential ‘there,’ quantified noun phrases, 624
 Exophora, 352
 Expectation theory, 413–414
 Experience *see a priori/a posteriori* distinction;
 Empiricism; Knowledge; Perception
 Experimental Analysis of Behavior, 31, 32
 Explanation, in language, 390
 Explication(s), 138, 182, 334
 Explicatures, 690, 692
 relevance theory, 651
 Exploiting, cooperative principle, 101
 Explosion, negation, 510, 511, 559
 Expositives, speech acts, 709
 Expression(s), 368, 369
 character and content, 44–46
 co-designative, 125, 126
 interpreted logical forms, 339–340
 speaker-meaning, 214–216, 333–334, 506, 528
 Expressive(s), 499–500
 speech acts, 709
 Expressive power of language, 217–219
 affective meanings, 217
 cardinal numerals, 218
 clause chaining, 219
 clauses, 218
 conceptual meanings, 217
 crosslinguistic synonymy, 217–218
 definition, 217
 extent, 217
 form–meaning relationship, 217
 ineffability, 217
 intonation, 218, 219
 linguistic types, 217, 218
 morphemes, 218, 219
 phonological features, 218
 polysynthetic languages, 219
 ‘snow words’, 217
 sound symbolism, 218
 syntax, 219
 synthetic languages, 219
 tonal languages, 219
 words, 218
 Expressivism, 532–533, 542, 635, 636–637
 Extended Standard Theory (EST), history/
 development, 790, 791
 Extension
 analycity, 15–16
 circumstances of evaluation, 45
 context, 125
 definition, 139
 multiple, 525–526
 two-dimensional logic, 794
 workers in, Frege, 125, 336–337, 682
 Extensionality, 219–222
 definition, 219, 221
 Montague semantics, 489
 Externalism
 causal theories of reference, 42
 content, 222–225
 E-language, 179, 180
 epistemology, 224
 meaning, 190, 641
 natural kind terms, 503
 reference, 637, 643
 self-knowledge, 190, 525, 537
 see also Object-dependence
- F**
 Facts
 brute *vs.* institutional, 703, 704
 denotation, 457
 intentional, 336, 703–704
 metaphysics, 456–457
 moral, 541, 542
 semantic, 336, 662–663
 slingshot argument, 456–457
 substitution *salva veritate*, 457–458, 464
 tensed, 742
 truth, 777, 780
 workers in
 Russell, 402–403, 777–778
 Wittgenstein, 403, 777–778
 Fallibilism, 560
 Falsificationism, 373, 373f, 560
 Familiarity, 132
 Farrell, F B, 448
 Fauconnier, G, 201
 Felicity conditions
 speech acts, 706, 708
 Feyerabend, P, 373, 560
 Feys, R, 467
 FI *see* Full Interpretation (FI)
 Fiction
 discourse, 227–230
 function, 227
 possible worlds *see* Possible worlds
 Field, H, 142, 636, 779
 Fields of inquiry, 382
 Figurative language, 230–242
 definition, 230
 examples of traditional figurative language, 231
 literal *vs.* figurative language, 237
 selected review of theories of metaphor, 231
 Aristotle, 231
 Black, 234
 Groupe μ , 234
 Johnson, 235
 Lakoff, 235
 Langacker, 237
 Richards, 233
 Ricoeur, 235
 Vico, 232
 semiosis and the signifying order, 238
 theories of metaphor, 231
 the uses of metaphor, 239
 anatomy, 241
 L2 pedagogy, 241
 linguistics, 240
 physics, 241
 theory-constitutive metaphors, 240
 File change semantics, 173, 176
 Fillmore, Charles J
 fields of work
 generative grammar, 788
 transformational grammar, 757
 First-order logic, 163, 221, 279–281
 modal operators, 469
 Firth, John Rupert, 785
 fields of work
 linguistic anthropology, 475–476
 see also London School
 Florentine *see* Italian
 Flouting, 421–424
 cooperative principle, 101
 definition, 423
 Fodor, J A, 204–205, 480–481, 562
 fields of work
 causal theory of meaning, 42
 cognitivism, 51
 frame problem, 267
 innate concepts, 328
 Language of Thought, 23, 53, 334–335, 443
 mind, 605–606
 misrepresentation, 657
 naturalization of content, 508
 probabilistic theory of concepts, 79
 Lepore, E, association with, 297, 508
 Fogelin, R, 508
 Foley, W A, 271–272
 Force, illocutionary, 497, 500
- Form *see* Logical form
 Formalism/formalist linguistics, 258–265
 autonomy, 258, 263
 formalist view, 258
 mentalist view, 258
 Chomskyan linguistics, 258, 259
 competence, 259, 260
 crosslinguistic perspective, 260
 data, 260
 E-language, 262
 government and binding theory, 261
 grammar, 260
 grammaticality judgments, 259–260
 I-language, 259, 260, 264
 interlanguage, 264
 L2 acquisition, 264
 language acquisition, 260, 264
 language change, 264
 language faculty, 260, 264
 language use, 263–264
 minimalist program, 261
 performance, 259–260
 principles and parameters model, 261, 264
 research program, 259, 261f
 standard theory (ST), 259, 260, 261, 263
 Universal Grammar (UG), 260
 vs. linguistic theory, 259
 definition, 258
 empiricism, 259
 formalist approaches, 262
 context-free grammar (CFG), 263
 Generalized Phrase Structure Grammar (GPSG), 263
 generative grammar, 262
 Montague grammar, 263
 semantics, 263
 syntax, 263
 functionalist linguistics, 259, 263–264
 language use, 263
 autonomy, 263
 communication, 263–264
 language acquisition, 264
 language change, 264
 principles and parameters, 264
 relevance theory, 263–264
 mentalist approaches, 261–262
 HPSG, 262, 263–264
 Lexical Functional Grammar (LFG) (LFG), 261–262, 261f
 psychological reality, 261–262
 research program, 259
 sociolinguistics, 259
see also Autonomy; Chomsky, Noam;
 E-language *vs.* I-language; Functionalism;
 Grammar; HPSG (Head driven Phrase
 Structure Grammar); Lexical Functional
 Grammar (LFG); Minimalism; Montague,
 Richard; Politeness; Pragmatics; Principles
 and Parameters Model
 Formal semantics, 97, 199, 243–258, 583
 definition, 243
 see also Compositionality; Implicature; Logic;
 Monotonicity; Presupposition; Quantifiers
 Formal surrogates, 171
 Foundationalism, epistemic, 183
 Fowler, Henry Watson, 149
 Frame(s)
 conceptual, 503, 690–691
 Frame problem, 265–267
 nonmonotonic logic, 524–525
 Yale shooting problem, 526
 Frames of Reference, 564
 France
 Académie Française, 113
 history of linguistics, 784
 Frege, Gottlob, 220, 514, 516
 fields of work
 analycity, 96–97
 character/content distinction, 45–46
 compositionality, 73, 680
 context principle, 88, 89–90, 295–296
 deflationism, 142
 denial, 512–513
 descriptivism, 41, 609–610, 612, 638, 640

- Frege, Gottlob (*continued*)
 existence, 212, 717
 ideational theories of meaning, 299, 301
 intension and extension, 125, 336–337, 682
 language and social construction, 702
 logic and language, 410
 mediated reference theory, 157, 433, 435,
 610, 696, 697, 798
 mentalist semantics, 688, 693
 object dependence theory, 536
 predication, 586
 principle of compositionality, 487–488
 propositions and sentences, 28, 124, 464, 620
 quantification, 716
 redundancy theory of truth, 779
 semantic value, 680
 singular terms, 798, 799
 thought, 124
 use theory of meaning, 801
see also Ideal language philosophy; Sense and
 reference
- Fregeanism
 belief, 191–192
 empty names, 186
 Frege's problem, 159–160, 336, 433, 614, 658,
 735, 772–773
- French, 411
- Fried, C, 414
- Full Interpretation (FI), 599
- Function
 properties, 547
 semantic relations, 427–428
see also Relational grammar; Systemic
 functional grammar
- Functional Generative Description (FGD)
 history/development, 791–792
- Functional grammar
 history/development, 791–792
see also Adjectives
- Functional Independence Principle, 690
- Functionalism, 268–275
 basic tenets, 268
 communication as primary function, 268,
 270, 271–272, 272–273
 integrative functionalism, 268–269
 motivation, 268–269, 270,
 271–272, 272–273
 nonautonomy of syntax, 269, 271, 272, 273
 further features, 269, 272
 flexibility of meaning and structure, 270
 grammar, 269, 270, 271, 273, 274
 language acquisition, 270, 271, 272, 273
 typological orientation, 270, 271, 272, 273
 use of authentic textual data, 269
- linguistic theories, 268
 emergentism, 274
 European/North American
 functionalism, 268
 functional grammar, 270
 role and reference grammar (RRG), 271
 systemic functional grammar (SFG), 272
 usage-based functionalist–cognitive
 models, 274
vs. formalist theories, 268, 269
 West Coast functionalism, 273
- mind, 296
 origins/development, 791
see also Cognitive grammar; Minimalism;
 Principles and Parameters Model; Role and
 reference grammar (RRG); Systemic
 functional grammar
- Functional structure–constituent structure relation
see C-structure
- Function argument
 logic, 400
- Function–argument structural descriptions, 352
- Fundamentals, linguistic reality, 368–371
- Futures, possible, 739
- Future tense, 276–278
 epistemology, 276
 grammaticalization, 276
 immediacy, 277
 inflectional marking, 277
 periphrastic marking *vs.*, 277
- marking devices, 276–277
 periphrastic marking, inflectional marking
vs., 277
 philosophy of time, 276
 ‘pragmatic strengthening’, 277
 prediction-based *vs.* intention-based, 277
 ‘semantic bleaching’, 277
 temporal distance, 277
 typological surveys, 277
- Futurity, 636–637, 741
- Fuzzy logic, 513
see also Vagueness
- G**
- Gale, R, 741
- Game-Theoretical Semantics, 76,
 279–282
- Game theory, 532
- Gapped bivalent propositional calculus *see* GBPC
- Gappy Proposition View, 187, 188
- Garden-path theory
 pragmatics, 319
- Gazdar, G, 263
- GBPC, 592
- Geach, Peter Thomas, 169, 302, 303
- Gender studies
 cooperative principle, 105
- Generalized conversational implicatures,
 relevance theory, 651
- Generalized Phrase Structure Grammar
 (GPSG), 263
 history/development, 790–791
- Generalized quantifiers, 36, 164,
 482–485, 494–495
 definition, 483
 scope, 627
- Generative grammar, 282–284, 347, 381
 formal grammar, 282
 analysis of natural language, 282–283
 components, 282
 recursion, 282–283
 rewriting rules, 282
 impact of, 789, 790
 mentalist (Chomskyan) approach, 282, 283
 E-language, 283
 I-language, 283
 Initial State, 283–284
 Innateness Hypothesis, 283–284
 internalism, 283
 language acquisition, 283–284
 language faculty, 283–284
 Steady State, 283–284
 Universal Grammar, 283–284
 origins/development, 787
 standard theory, 788
- Principles and Parameters Model, 596–605
see also Binding theory; Minimalism; X-bar
 theory
- workers in
 Chomsky, Noam, 347, 787
 Fillmore, Charles J, 788
 Katz, Jerrold J, 788
 Klima, Edward, 788
see also Chomsky, Noam; Cognitive science;
 Principles and Parameters Model
- Generative lexicon, 579
see also Compositionality; Lexical conceptual
 structure; Lexical semantics; Syntax,
 semantics interface; Thematics, relation
- Generative phonology
 origins/development, 788
- Generative semantics, 284–289, 760
 analyses, 287
 controversies, 285
 history of, 287
 origins/development, 790
 reserach conduct, 286
 workers in
 Lakoff, George, 790
 McCawley, James D, 790
 Postal, Paul M, 790
 Ross, John R, 790
- Generativism
 history of, 477
 workers in
 Chomsky, Noam, 477
 Skinner, B F, 477
- Generic categorization, 349
- Generic conception of language,
 563–564, 564–565
- Generic quantifier, 193–194
- Generic reference, 289–293
 definition, 289
 forms of, 289
 bare plurals, 290
 classification systems, 289
 count terms, 290
 cross-languages, 291
 definite noun phrases, 290
 generics (habituals), 291
 kind-referring phrases, 289–290
 mass terms, 290
 singular indefinite articles, 290–291
 taxonomic reading, 290
 undetermined mass expressions, 290
 theory of, 291
 bare plurals (**English**), 292
 descriptions, 291
 individual-level predicates, 291–292
 kind-level predicates, 291–292
 quantification, 291
 stage-level predicates, 291–292
 undetermined mass expressions, 290
see also Generics (habituals)
- Generics (habituals), 291, 533–533
see also Generic reference
- Geneva School
 history, 783
- Gentner, D, 200
- Geometry
–a priori knowledge, 1
- Germinate consonants, quantity *see* Quantity
- Geurts, B, 518–519
- Gilliéron, Jules
 fields of work
 linguistic geography, 475
- Ginzburg, J, 687
- Givenness hierarchy, 136
- Givón, Talmy, 273, 516
- GL *see* Generative lexicon
- God
 ineffability, 365, 366
 Gödel, K, 8–9, 404, 405, 408, 461
- Goldman, A, 507
- Goodman, N, nominalism, 370, 523
- Good manners *see* Politeness
- Gordon, D, syntax–pragmatics interface, 718
- Gorgias, *On What Is Not*, 569–570
- Gould, Stephen J, 205
- Government-binding theory, 261, 382
see also Principles and Parameters Model;
 Transformational grammar
- GPSG *see* Generalized Phrase Structure Grammar
 (GPSG)
- Gradable comparatives, 68–69
 adjectives, 424
- Gradable complementarity *see* Antonyms/
 antonymy
- Gradable contrariety *see* Antonyms/antonymy
- Grady, J, 203
- Grammar, 368
 architecture, 21–25
 categorical *see* Categorical grammar
 cognitive grammar *see* Cognitive grammar
 comparative *see* Comparative grammar;
 Comparative philology
 competence theory
 performance *vs.*, 24–25
 Context-free *see* Context-free Grammar (CFG)
 cooperative principle, 104
 cross-laguage variations, 426–427
 design, 25
 generative *see* Generative grammar
 HPSG *see* HPSG (Head driven Phrase Structure
 Grammar)
 levels of representation, 411, 412

- logical form, 7, 153, 411
 recursion, 566
 reference, 434
 relational *see* Relational grammar
 space grammar *see* Cognitive grammar
 surface, 153–154, 229, 402, 799
 systemic-functional grammar *see* Systemic functional grammar (SFG)
 tacit knowledge, 734
 theories of
 relevance theory, 654–655
 syntactic (generative) *see* Generative grammar
 Universal Grammar *see* Universal Grammar
 transformational *see* Transformational Grammar
 Tree-Adjoining *see* Tree-adjoining grammar (TAG)
 unification-based, 23–24
 universal grammar *see* Universal grammar
 workers in
 Montague, 547, 681, 697
see also Context-free grammar; Generalized Phrase Structure Grammar; Generative grammar; Lexical Functional Grammar (LFG); Montague grammar; Role and reference grammar; Systemic functional grammar; Universal grammar
- Grammar inclusive theory, syntax-pragmatics interface, 726
- Grammaticality judgments, 118
 criticisms of, 118–119
 ‘expert’ judgments, 118
 judgements *vs.* own usage, 118
 possible *vs.* impossible utterances, 118
 scaling, 118
- Grammaticalization, 150
 future tense, 276
- Grammatical meaning, 293–294
 concrete *vs.* relational concepts, 293
 definition, 293
 inflectional markers, 294
 lexical meaning *vs.*, 293–294
 prosody, 294
 Sapir, Edward, 293
 subjectivity *vs.* objectivity, 293
 word order, 294
- Grammaticized verbs, evidentiality *see* Evidentiality
- Greenbaum, Sidney, 150
- Greenberg, Joseph H
 language types, 787
- Grelling-Nelson Paradox, 557–558
- Grice, Herbert Paul, 247–248, 255
 fields of work
 causal theory of meaning, 43
 convention, 99, 773–774
 Cooperative Principle, 427, 688–689
 expression *vs.* speaker meaning, 216, 333–334, 646, 773–774
 intentions, 334, 415, 773
 irony, 341
 maxims, 447, 532, 647–648, 688–689
 mental representation, 656
 metaphor, 450
 new philosophy of language, 53, 56
 ordinary language, 8, 49, 50, 551
 politeness *see* Politeness
 public language, 750–751
 relevance theory, 439
 speech-act theory, 478
see also Implicature
 pragmatics, 615–616, 688–689, 690, 691, 792–793
 publications
 Studies in the Way of Words, 100–101
 Strawson, Peter, association with, 16
 Grice’s circle, 689
 Grice’s Cooperative principle (of conversation), 429
- Grimm, Jacob Ludwig Carl
 fields of work
 comparative philology, 473
- Groenendijk, J A G, 176
 Stokhof, M, association with, 21, 683–684, 686–687
- Grotius, Hugo, origin of language, 554
- Grounding, 87
see also Common ground
- GTS (game-theoretical semantics) *see* Game-theoretical semantics
- Gualmini, A, co-workers and associated workers
- Crain, S, 332
- Pietroski, P, 332
- Guguyimidjir
 Frames of Reference, 564–565
- Guillaume, Gustave, 785
- Gundel, J K, 136
- ## H
- Haack, S, 774, 776–777
- Habituals *see* Generics (habituals)
- Halliday, Michael A K, 272–273, 785
 fields of work
 linguistic anthropology, 475–476
 predication, 586
- Hamblin, C, 686
- Handbook of American Indian Languages (Boas), 475
- Hanks, S, association with McDermott, D, 266
- Hapology
 Optimality Theory, in morphology *see* Optimality Theory
- Hare, R M, 8
- Harm, lying, honesty and promising, 413–415
- Harman, G, 17, 336, 442
- Harnish, R M, association with Bach, K, 334
- Harris, Zellig S, 786–787
 fields of work
 distributionalism, 477
 transformational grammar, 753
- Hauser, M D, 204–205
- Hawkins, John A, 792
- Hayes, P, 267
 McCarthy, J, association with, 265
- Head Directionality Principle, 332
- Head-Driven Phrase Structure Grammar (HPSG) *see* HPSG (Head driven Phrase Structure Grammar)
- ## Hebrew
- language revival, 113–114
- Hegel, G W F, 702
- Heidegger, Martin, 365, 366, 367, 812
- Heim, Irene, 20, 132–133, 176
- Kratzer, A, association with, 681
- Helmholtz, Hermann Ludwig Ferdinand von, 1–2
- Hempel, C G, 33, 295, 812
- Henken quantifier sentences, 279–281
- Heraclitus, 569
- Herder, Johann Gottfried, 554/
 fields of work
 origin of language, 554
- Hesiod, 569
- Hidden argument place, 193
- Higginbotham, J, 336, 338, 339, 529
- Higher order groups, 576
- Hindi
 use of
 number of speakers, 218
- Hintikka, Jaakko, 685
 Sandu, G, association with, 447
- Hjelmlev, Louis Trolle, 784
- Hobbes, Thomas, 299, 446, 553/
 origin of language, 554
- Hockett, Charles Francis, 786–787
- Hodges, W, 74, 77
- Holism, 295–297
 compositionality, 71–72, 74–75
 confirmation, 14, 34, 183–184, 295
 context principle, 88, 90–91
 epistemic, 14, 190
see also Semantic holism
- Homer, 569
- Homonymy, 577–580
 definition, 577
- Honesty, 413–415
- Hopper, P J, 274
- Horn, L R, 517
- Horned Man Paradox, 591
- Hornian pragmatics *see* Neo-Gricean pragmatics
- Hornsby, J, 4
- Hornstein, N, association with Antony, L, 733
- Horwich, P, 76, 142, 143, 803
- How to do things with words (Austin), 478
- HPSG (Head driven Phrase Structure Grammar), 23, 24, 262, 263–264
 syntax-pragmatics interface, 720, 721f, 722f
- Hull, C, 31
- Humanity principle, 633–634
- Humberstone, L, association with Davies, M, 796
- Hume, David
 empiricism, 183, 184
 existence, 212
 ideational theory of meaning, 299
 identity, 303–304
 positivism, 810
- Humor
 irony, link with, 342
 relevance theory, 654–655
- Humphrys, J, 150
- Hurford, J R, 207
- Hutcheson, Francis
 moral sense theory, 183
- Hyperonymy, 349
 entailment relations, 256
- Hypertensionality, 460, 461, 462, 463, 464–465
- Hyponymy, 349, 359
- Hypotheticals, 497, 498, 499, 499t, 500
- Hypothetico-deductive method, 560
- ## I
- Icelandic
 case relations, 599–600
- Idealism
 mind and body, 618
 realism, 634–635, 636–637
- Idealization, 375
 linguistics as science, 395
see also Psycholinguistics
- Ideal language philosophy, 49, 410
 reference-based theory, 47, 49, 53
see also Frege, Gottlob; Quine, Willard van Ormand; Russell, Bertrand
- Ideas
 innate *see* Innate ideas
 simple/complex, 300
- Ideational theories of meaning, 299–301, 433, 638, 771
- Identifiable morphemes, in isolating language *see* Chinese
- Identity, 107, 220
 modal logic, 469
 PTQ, 492–493
see also Partial identity
- Identity, philosophical aspects, 302–304
 numerical, 302, 305, 306
 relative/absolute, 302
 sameness, 302
 synchronic, 303–304
see also Misidentification
- Identity sentences, 610, 696
- Identity theory of mind, 798
- Idiolect
 I-language, 179, 180, 339
- Idiom(s), 499–500
- Idiom model
 indirect speech acts, 711
- IEM *see* Immunity to misidentification error
- IF (independence-friendly) logic, 279–281
- If-then sentences *see* Conditionals
- I-language (internalized), 561, 563–564
 E-language *vs.* *see* E-language *vs.* I-language
- ILFs *see* Interpreted logical forms
- Illocution
 illocutionary act, 333–334, 437, 548, 689
 illocutionary force, 497, 500
 illocutionary point, 497–498, 499–500

- Illocution (*continued*)
 intention, 497–498, 500
 primary, 497
- Illocutionary force indicating device (IFID),
 speech acts, 707–708
- Illocutionary speech acts, 707
 speech acts, 707
- Imagery, mental, 300, 365
- Image schemas
 Lakoff, G & Johnson, M, 236
- Immediacy
 future tense, 277
- Immunity to misidentification error, 305–308
- Imperative, 496–497, 498–499, 499*t*
- Imperative mood, 497
- Implicature, 50, 99, 247–248, 254, 318, 647–648
 assertions, 29–30
 generalized conversational, 215, 529,
 689, 691–692
 relevance theory, 651
 Gricean types, 688–689, 691, 692, 694–695
 intention, 532
 metaphors, 447
 negation, 517
 particularized, 528
 relevance theory, 651
 speech-act theory, 8
- Implicatures, 529, 691
- Implicit variables, 192–193
- Inclusion
 semantic relations, 427–428
- Indefinite(s), 131–138, 677–678
 definition, 131
see also Pronouns
- Independence-friendly (IF) logic, 279–281
- Indeterminacy
 semantic, 52–53, 92, 322–324
 comp, 72, 632–633
 Quine, 52–53, 322
 of translation, 67, 322, 436, 437, 508, 631, 772
- Indeterminate future, 739
- Indexicality, 317
 philosophical aspects, 324–327
- Indexicals
 character/content distinction, 44, 159, 216, 795
 context, 44, 324, 326, 691
 double- *vs.* single-indexing, 45
 essential, 191–192
 hidden, 615, 692–693
 meaning, 617–618
 object-dependent thoughts, 535–536, 537
 pure, 325
 reference, 41, 157, 159, 326, 637, 640–641, 643
 temporal, 741, 742
- Indicative conditionals, 81
- Indicative mood, 497, 498, 685
- Indirect (mediated) reference, 157, 159
 workers in
 Frege, 433, 435, 614–615, 697
 Russell, 615
- Indirect speech acts, 315, 709
 analysis, 710
 conventionalization, 710
 conversational postulates, 711
 idiom model, 711
 literal force hypothesis, 710
 pragmatic acts *see* Pragmatic acts
- Indiscernibility of identicals, 302, 303
- Individualism
 E-language/I-language distinction, 179
 semantic holism, 296
- Individual-level predicates, 193
 generic reference, 291–292
- Induction, 521
- Inductivism, 371, 372*f*
- Ineffability, 365
- Infant(s)
 attention in
 joint *see* Joint attention
 nonlinguistic goals, 429
see also Children
- Infectional markers
 grammatical meaning, 294
 lexical morphemes *vs.*, 294–294
- Infelicity *see* Presupposition
- Inference, 531–532
 default rules, 525, 526
 detachment, 5
 implicatures, 529
 mental logic, 443
 models, relevance theory, 649
 modus ponens, 532–533, 776–777
 nonmonotonic, 524–527
 representation, 334–335, 658
 rules, 407–408, 408–409
 testimony, 745
 workers in
 Boole, 399
 Frege, 400, 410
 Leibniz, 398
 Russell, 183, 410
see also Projection problem
- Inferring, 317
- Infinity, 366, 367
- Inflectional phrase (IP)
 definition, 388
see also Sentence
- Informational independence, 279–281
- Informative intention, relevance theory,
 649–650
- Innate ideas, 184, 327–330, 736–737
 concept nativism, 80, 183, 327, 329–330, 443
- Innate knowledge, 330–333
 cognitive linguistics, 51–52
 content, 331–332
 language acquisition, 331, 332, 566, 567–568
 representational/computational
 modules, 331–332
 universal grammar, 179, 567
- Innateness Hypothesis, 283–284, 561–562
- Instantiation, 212, 545, 563–564
 abstract objects, 545–546, 546*f*
- Instrumentalism, 560–561, 633, 636–637
- Instruments (of an action), adverbial phrases, 193
- Integrative functionalism, 268–269
- Intellect, active/receptive, 521
- Intelligence planning theory, artificial, speech acts
see Speech acts
- Intelligibility
 limits, 365–367
- Intensionality-compositionality merger, 128,
 129, 130
- Intentional circle argument, 34
- Intentionality/intention, 128, 219–222
 action, 3, 97
 advertising, 335
 analycity, 16
 collective, 703, 704–705
 communicative *see* Intentionality/intentional
 communication
 compositionality, 334
 context, 125
 definition, 219
 donkey anaphora, 169
 I-language, 179, 180
 illocutionary, 497–498, 500
 as indirect referent, 614–615
 infants, 197
 intensional stance, 633
 intention-reading skills, 197, 198*t*
 interpreted logical forms, 339
 language and social construction, 703
 language and thought, 333
 meaning, 333–336, 506, 773
 mind, 749–750, 751
 modal logic, 470
 Montague semantics, 489
 normativity, 532
 promising, 414–415
 reflexive, 334
 representational, 334–335, 613, 656, 658
 semantic value, 682
 two-dimensional, 794, 794*f*, 796–797
 workers in, Frege, 125, 336–337
- Intentionality/intentional communication, 64–65,
 98–99, 215, 216, 334, 335, 693
- Interlanguage, 264
see also Competence
- Inter-language differences *see* Crosslinguistic
 studies/variation
- Intermediate representation level, 174
- Internalism, 23, 179, 222, 561, 563, 566
 demonstratives, 223
 mentalism, 180, 796–797
 radical internalism, 563
 reference, 223, 641, 643
see also E-language *vs.* I-language
- Interpretation, 408–409, 613–614, 616, 688
 radical, 631, 772–773
see also Translation
- Interpretationalism, 323–324
- Interpreted logical forms, 336–340
 Hard Name Puzzle, 337
- Interpretive semantics
 transformational grammar, 763
- Interrogative
 clauses, 496–497, 498, 498*t*
 interrogative mood, 497, 685
 metaseantics
 epistemic-imperative approach, 685
 force/radical approach, 685
 question-as-answer approach, 686, 687
 truth-conditional semantics, 685–687
- Intersective determiners, 625, 626
see also Existential quantifier
- Interval semantics, 739
- Intransitive verbs *see* Verb(s)
- Intrusion, pragmatic, 689–690, 691–692, 693
- Intuition, 374, 375, 376, 377, 562–563
see also Inference
- Intuitionism, 405, 635, 812–812
- Inventories, Optimality Theory *see* Optimality
 Theory
- IR *see* Indirect (mediated) reference
- Irish Gaelic, 114
- Irony, 341–343, 528
 definition, 341
 direct access theories, 341–342
 Grice, H Paul, 341
 humor, link with, 342
 markers, 342
 mention theory, 341
 as negation, 341
 psycholinguistics, 342
 relevance theory *see* Relevance theory
 sarcasm *vs.*, 341
 sociolinguistics, 342
- Island constraints, 601, 767
- Israel
 Academy of the Hebrew Language, 113–114
- Italian**
 binding constraints, 601
- Itkonen, E, 377, 563
- J**
- Jackendoff, Ray S, 53, 181, 204–205, 206, 207,
 208, 209*f*, 329, 349, 354, 356, 362, 693, 694
- Jackson, F, 8–9, 796–797
- Jakobson, Roman, 783
- James, William, 31, 183–184, 778
- Jespersen, Otto, 417–418
- Johnson, Mark, 202
 fields of work
 figurative language, semiotics of, 235
 Lakoff, G, association with, 445–446
- Johnson, Samuel, 147–148
- Joint attention
 lexical meaning, 427, 431–432
- Joint experience *see* Common ground
- Jones, Daniel, 785
- Jones, William
 fields of work
 comparative philology, 473
- K**
- K (system of modal logic), 466
- Kamp, Hans, 20, 45, 176, 682–683, 739
- Ryle, U, association with, 616

- Kanazawa, M, 21
 Kandel, E, 35
 Kant, Immanuel
 fields of work
 ~a priori knowledge, 1, 2, 3, 11
 analytic/synthetic distinction, 1, 3, 13, 96–97
 existence, 212
 judgment, 13, 90–91
 lying, 415
 promising, 414
 syncategoremata, 716, 717
 truth as necessary, *~a priori* and analytic, 10
 ultimate reality, 366
 Kaplan, D, 171
 fields of work
 antidescriptivism, 640
 causal-historical theory of reference, 41, 159, 641–642
 demonstratives, 648, 682–683, 795
 direct reference, 157, 158, 159, 615–616
 sense and reference, 697, 698
 two-dimensional semantics, 795, 796
 see also Character/content distinction
 Karmiloff-Smith, A, 562
 Karttunen, L, 686–687
 Kasher, A, 688
 Katz, Jerrold J, 16–17, 74–75, 347–348, 562
 fields of work
 generative grammar, 788
 Katz-Postal Hypothesis, 760
 Kind-level predicates, generic reference, 291–292
 Kind-referring phrases, generic reference, 289–290
 King, J, 198, 325–326
 King, J C, 20
 Kirkham, R L, 142
 Kitcher, P, 507
 Kittay, E F, 448
 Kleene, S C, 593
 Klima, Edward S
 fields of work
 generative grammar, 788
 Knowledge
 ~a priori, 1–3, 11, 15, 17, 46, 189
 Ayer, 14, 15, 96–97
 Quine, 2, 11, 12
 assertions, 29
 causal thesis, 183
 innate, 51–52
 language, 180, 189–191
 of language, 384
 normative thesis, 183
 a posteriori, 1, 11, 183
 of self, 190, 224, 525, 537
 tacit knowledge, 733–737
 tacit, 180, 504
 propositional attitudes, 733, 735
 of universal, 523
 of world, 224
 see also Innate knowledge
 Kornblith, H, 507
 Kratzer, A, 82–83, 84
 fields of work
 hidden argument place, 193
 Heim, I, association with, 681
 Kretzmann, N, 714
 Krifka, M, 194–195
 Kripke, Saul
 fields of work
 analytic/synthetic distinction, 3
 causal theory of reference, 41–42, 159, 610, 641–642
 definition, 17
 descriptivism, 502, 611, 640–641
 determinacy of meaning, 322
 direct descriptions, 154
 direct reference theory, 615–616
 externalism, 503, 637
 meaning skepticism, 437–438, 662–663
 natural kind terms, 503, 612
 necessary *a posteriori* and contingent *~a priori*, 796
 possible worlds semantics, 8–9, 580
 proper names, 771
 propositions as abstract entities, 618
 public language, 750–751
 rigid designation, 158–159, 304, 660
 rule-following, 661
 sense and reference, 158, 435–436, 697, 698
 speaker-meaning *vs.* expression meaning, 215, 646
 truth predicates, 559
 Kuhn, T, 373
L
 Labov, William
 fields of work
 sociolinguistics, 792
 Lacan's twin doors, 675f
 Lakatos I, 373
 Lakoff, George, 185, 202
 fields of work
 figurative language, semiotics of, 235
 generative semantics, 790
 syntax-pragmatics interface, 718
 Johnson, M, association with, 445–446
 Lamb, Sydney M, 787
 Lambda-calculus *see* Lgr-calculus
 Langacker, Ronald W, 200
 fields of work
 figurative language, semiotics of, 237
 Language, 735–736
 acquisition *see* Language acquisition
 cognitive development and *see* Language-thought relationship
 communication, 66, 299
 competence models, 22
 development *see* Language acquisition
 epistemology *see* Epistemology
 evolution *see* Origin and evolution of language
 explanation in, 390
 figurative *see* Metaphor
 formal, 407, 558, 775
 inter-language differences *see* Crosslinguistic studies/variation
 knowledge of, 384
 descriptions, 388
 learning *see* Language learning
 logic *see* Logic
 magic, 73–74
 meaning of, 383
 non-literal, 367
 non-standard use, 527–530
 origin of *see* Origin and evolution of language
 performance models, 22
 possible, 97, 98
 Principles and Parameters model, 24
 processing *see* Language processing
 production *see* Language production
 public nature, 606, 656, 750–751, 771
 semantically closed, 558
 standardized use, 527, 529
 as system of knowledge, 566
 variation
 communal common ground, 86
 see also Conventions; Externalism; Ideal language philosophy; Internalism; Metalanguages; Object language; Ordinary language philosophy
 Language acquisition, 347, 352
 bilingualism *see* Bilingualism
 constraints, 430
 conventions, 99
 crosslinguistic differences *see* Crosslinguistic studies/variation
 empiricism, 185, 508
 errors, 331
 functionalist view, 270, 271, 272, 273
 innateness, 331, 332
 language as prior to thought, 751
 naturalism, 561–562
 pragmatics *see* Pragmatics development
 rules, 567
 syntax *see* Syntax, development
 vocabulary/meaning development, 426–432
 see also Lexical acquisition; Semantic development
 see also Applied linguistics; Infant language acquisition; Language development; Language learning; Learnability
 Language and social construction, 702–705
 Language as an object of study, 345–348
 E-language, 347
 I-language, 347
 language as a mental organ, 346
 generative grammar, 347
 generative linguistics, 346
 language acquisition, 347
 psychology, 347
 language as an abstract object, 347
 language as a social fact, 345
 langue *vs.* parole, 345–346
 synchronic perspective, 345
 language as behavior, 346
 behaviorism, 346
 logical positivism, 346
 properties of language, 345
 Language faculty, 25, 180, 181, 260, 264, 283–284
 as innate, 331, 567–568
 tacit knowledge, 734, 735
 Language learning, 35, 73, 185, 567
 Language of Thought, 23, 53, 334–335, 523, 656
 see also Associationism and connectionism; Cognitive science; Fodor, J A; Mind, philosophy of
 Language processing, 23, 53, 99, 728–729
 pragmatic *see* Pragmatics
 Language production, 22, 567, 613–614, 616
 Language revival, 113
 Irish Gaelic, 114
 Modern Hebrew, 113–114
 modern vocabulary, 113–114
 writing systems, 114
 Language-thought relationship, 53–54, 396, 613–614
 behaviorism, 222
 intention, 333
 Locke, 299, 333
 philosophical issues, 715–716, 747–752
 relative priority, 300, 548
 see also Mentalese
 Langue, 345–346
Langue-parole dichotomy, Saussurian linguistics, 474
 Larson, R K, 336, 339–340
 co-workers and associated workers
 Ludlow, P, 338–339, 339–340
 Segal, G, 54, 339–340
 Lasnik, H, 768
 Lass, Roger, 562–563
 Lattice-theory, 37
 Laurence, S, 98–99
 Law of excluded middle, 279–281
 Laycock, H, 417
 Lazaraton, A, 376
 LCS *see* Lexical conceptual structure (LCS)
 Learnability, 73, 353, 355, 688
 see also Language acquisition
 Leddy, T, 448
 Lees, Robert B, 788
 Leibniz, Gottfried Wilhelm
 fields of work
 logic, 398, 399, 804
 Leibniz's Law, 302, 303–304
 Le Poidevin, R, 741
 Lepore, E, co-workers and associated workers
 Fodor, J, 297, 508
 Loewer, B, 339, 615
 Levin, S R, 446
 Levinson, Stephen C, 215, 529, 564, 689, 690, 691–692
 Levinsonian pragmatics *see* Neo-Gricean pragmatics
 Lewis, David, 85, 468
 conventions, 180
 fields of work
 counterpart theory, 581
 language and social construction, 702–703
 mentalism, 688

- Lewis, David (*continued*)
 possible worlds, 580, 581–582
 reference, 155, 637, 638, 688
 semantic value, 681
- Lexical acquisition, 426–427
 coping strategies, 431
 cues
 syntactic, 429–430
 development of meaning, 426–432
 categorization, 428, 431
 class assignment, 428
 conceptual preferences, 431
 context, 427, 428
 contrast, 427, 429–430
 conventionality, 427, 429
 conversation and, 427
 semantic relations, 427–428
 syntax and, 429–430
 word learning experiments, 428, 430
- Lexical ambiguity
 disjunction, 248
- Lexical aspect *see* Aktionsart
- Lexical concepts, 196–197, 200, 200*f*, 201
- Lexical conceptual structure (LCS), 349–356
 definition, 349
see also Generative lexicon; Variable(s)
- Lexical conditions
 ambiguity, 577–580
 socially recognized function, 426
- Lexical decomposition, 355, 361
- Lexical density, reference *see* Reference
- Lexical Functional grammar, 23
- Lexical Functional Grammar (LFG), 24, 261–262, 261*f*
 history/development, 790–791
see also C-structure; Optimality Theory; X-bar theory
- Lexicalism, morphology *see* Morphology
- Lexicalist Hypothesis, 762
- Lexical meaning, grammatical meaning *vs.*, 293–294
- Lexical morphemes, inflectional markers *vs.*, 294–294
- Lexical pairs *see* Antonyms/antonymy
- Lexical semantics, 357–364, 692, 694
 ambiguity *see* Lexical ambiguity
 argument structure, 360
 classifications, 360
 definition, 357
 history, 357
see also Compositionality
- Lexicography, 112, 147–148, 243
- linguistic reality, 368
- Lexicon
 cross-language variation in
 lexicalization, 426–427
 development *see* Lexical acquisition
 generative *see* Generative lexicon
see also Formal semantics; Lexical conceptual structure (LCS); Word(s)
- Lexicon, mental, 692, 693
- LFG *see* Lexical Functional Grammar
- LGF *see* Grammar
- Liar Paradox, 511, 557
- Lieberman, P, 205
- Lightfoot, D, 207, 264
- Ligurian *see* Italian
- Linebarger, M C, 118
- Lingua characteristica universalis, 398
- Linguistic anthropology
 history, 475
 workers in
 Boas, Franz, 475
 Duponcaeu, Pierre Etienne, 475
 Firth, John Rupert, 475–476
 Halliday, Michael A K, 475–476
 Malinowski, Bronislaw, Kaspar, 475
- ‘Linguistic’ approach, anaphora, 144
- Linguistic Direction Principle, 691
- Linguistic dominance, 200, 200*f*
- Linguistic geography
 Gillieron, Jules, 475
 history of, 475
 Wenker, Georg, 475
- Linguistic knowledge, 562–563
 internalized knowledge, 563
 mutual knowledge, 563
 phonological knowledge, 563
- Linguistic relativity
 history of, 476
 Sapir, Edward, 476
- Linguistics, 378–383, 383–397
 applied
 see Applied linguistics
 cognitive *see* Cognitive linguistics
 computational
 see Computational linguistics
 corpus analysis *see* Corpus linguistics
 cross-cultural studies *see* Crosslinguistic studies/variation
 data, 379
 empiricism, 380
 linguistic rules, 380
 rationalism, 379–380
 statistics, 380
 definition, 368, 566
 empiricism, 379
 data, 380
 fields of inquiry, 382
 cycle, 382–383
 feature, 383
 morphology, 382
 phonetics, 382
 phonology, 382, 383
 pragmatics, 382
 semantics, 383
 syntax, 382, 383
 holistic theories, 381
 linguistic form, 378
 antilinguistic approach, 378
 autonomy, 378
 choice, 379
 constraints, 378–379, 381
 filters, 381–382
 formalist approach, 378
 functionalist approach, 378
 generative linguistics, 381
 government-binding theory, 382
 Optimality Theory, 382, 383
 production, 381
 relation to meaning, 378, 379
 status, 378
 systemic functional grammar, 379
 transformational rules, 378–379, 381
 as mathematics, 566, 568
 modular theories, 381
 naturalism, 381
 as psychology, 179, 566, 567
 rationalism, 379, 380
 data, 379–380
 idealization, 379–380
 realism, 381
 as science, 393, 566–568
 idealization, 395
 universals, 393–394
see also Autonomy; Data and evidence; Empiricism; Formalism/formalist linguistics; Functionalism; Generative grammar; Morphology; Phonology; Pragmatics; Psycholinguistics; Rationalism; Syntax; Systemic functional grammar
- Linguistics, history of
 19th Century, 472–480
 20th Century *vs.*, 781
 20th Century, 472–480, 780–794
 19th Century *vs.*, 781
 Copenhagen School, 784
 see also Copenhagen School
 France, 784
 Geneva School, 783
 London School *see* London School
 Prague School *see* Prague School
 USA, 785
 Aristotle, 25–27
 Chomsky *see* Chomsky, Noam
- Linguistics as a science, 371–377
 Bloomfield, 374, 376
- Chomsky, 374, 375
 cognitive science, 375
 competence, 375
 conversation analysis, 376
 data, 371, 375, 376
 deduction, 372
 descriptive linguistics, 376
 discovery procedures, 374
 evidence, 372
 falsificationism, 373, 373*f*
 history, 373–374
 hypotheses, 371–372, 372–373, 375, 376
 hypothetico-deductive model, 373
 idealization, 375
 induction, 372
 inductivism, 371, 372*f*
 intuition, 374, 375, 376, 377
 levels of adequacy, 375
 logical positivism, 371
 methods, 375–376
 Popperian model, 373, 373*f*, 375
 qualitative research, 376
 quantitative research, 376
 Saussure, 374
 sociological model, 373
 theory, 375
see also Chomsky, Noam; Data and evidence; Idealization; Language as an object of study; Structuralism
- Linguistic universals *see* Universals
- Literal force hypothesis, indirect speech acts, 710
- Literal meaning
 definite/indefinite descriptions, 154, 155
 implicature, 99
 metaphor, 447, 448, 528
 utterance *vs.* speaker meaning, 215, 333–334
- Literature
 relevance theory, 654–655
- Loar, B, 606
- Locative existentials, 134
- Locatives
 adverbial phrases, 193
- Locke, John, 183, 184, 436, 502, 553*f*
 fields of work
 descriptivism, 502
 ideational theory of meaning, 299, 301, 771
 innateness, 330
 metaphor, 446
 origin of language, 553
 skepticism, 744–745
 syncategoremata, 716
 thought and language, 299, 333, 548
- Locutionary speech acts, 437, 497–498, 548, 689, 707
- Locutions, 497–498
 modal, 581
- Loewer, B, association with Lepore, E, 339, 615
- Logic, 279, 398–406
 algebraic, 398
 autoepistemic, 525
 Boolean algebra, 36–39
 completeness theorem, 405, 408
 default, 525–526
 definition, 138–139
 deontic, 532
 entailment relation, 775
 extensionality, 221
 first-order, 408, 409, 411–412, 680–681
 formal logic, 407, 548
 formal semantics, 243–258
 function and argument, 400
 intensionality, 221
 intuitionism, 405, 408–409
 mental, 443
 multivalued (fuzzy), 513
 negation, 513–519
 nonmonotonic, 266, 524–527
 operators
 tense, 737–738
 polysemy, 578
 probabilistic, 526
 propositional logic, 399, 400–401
 second-order, 408–409
 situation calculus, 265

- two-dimensional modal, 794, 795
 vagueness, 808
 workers in
 Frege, 410
 Russell, 410
see also Artificial languages; Lgr-calculus;
 Dynamic predicate logic; First-order logic;
 Interpreted logical forms; Modal logic;
 Predicate logic; Propositional calculus;
 Second-order logic; Tense logic
- Logical consequence, 8–9, 336–340
 model-theoretic, 408, 409, 775
 proof-theoretic, 407–408
- Logical form (LF), 690
 different notions, 410–413
 interpreted, 336–340
 natural language, 410, 412
 surface grammar, 7, 153, 411, 583, 615
 workers in, Chomsky, 411, 412, 693, 694
- Logical Positivism, 217, 346, 371, 560, 561, 717
 analytic/synthetic distinctio, 2, 16–17
 behaviorism, 33
 meaning, 434–435, 772
see also Carnap, Rudolf; Verificationism
- Logicism
 Frege, 400
 Russell, 401–402
- Lombard *see* Italian
- Lombard, L B, 4
- London School, 785
 history, 785
see also Firth, John Rupert
- Long-distance wh-movement, 600–601
- Lowth, Robert, 147
- Ludlow, P, 336, 339–340
 Larson, R, association with, 338–339, 339–340
- Lust, B, 264
- Lying, 413–415
 non-informational theories, 415
- Lynch, J, 147–148
- M**
- Mackie, J L, 636
- Make-belief, 227, 228, 229
- Malapropisms, 214–215, 527
- Malay
 use of
 speaker numbers, 218
- Malcolm, N, 189
- Malinowski, Bronislaw Kaspar
 fields of work
 linguistic anthropology, 475
- Mandler, J M, 200
- Marcus, Ruth Barcan, 469
- Markerese, 358
- Marking, active/inactive *see* Active/inactive
 marking
- Marselius of Inghen, 522
- Martinet, André, 784
- Marxist language theories
 Saussurian linguistics, 668–669
- Mass nouns, 417
 count nouns, 417, 418, 419
 natural kind terms, 501
 semantic status, 417, 419–420
- Mass terms, 38, 489
 generic reference, 290
see also Numerals
- Material conditionals, 82
- Material implication, 466
- Materialism
 eliminative, 613
- Mates, B, 74
- Mathematics
 context principle, 89–90
 empiricism, 183
 linguistics, 566
 logi, 405, 406
 necessary truth, 12
- Mathesius, Vilem, 783
- Maxim(s), 421–424
 cooperative principle, 100
 criticisms, 422
 definition, 422
 drawbacks, 422
 manner maxim, 421
 quality maxim, 421
 quantity maxim, 421
 relation maxim, 421
 variety, 422
- Maxim of manner, 421
- Maxim of quality, 421
 implicature *see* Implicature
- Maxim of quantity, 421
- Maxim of relation, 421
- McCall, S, 741
- McCarthy, John
 Hayes, P, association with, 265
- McCawley, James D
 fields of work
 generative semantics, 790
- McConnell-Ginet, Sally, 5
- McDermott, D, 267
 Hanks, S, association with, 266
- McDowell, John, 158, 698
- McGinn, C, 685
- McTaggart, J E M, 741
- McTaggart's paradox, 742
- Meaning
 belief, 66–67
 causal theory, 43
 compositionality, 72, 90, 728, 771
 convention, 96
 default, 529, 689, 691–692
 deflationary theory, 143, 803
 direct reference theory, 143, 434, 641
 empiricism, 184, 299
 expression-centered, 215–216
 expression *vs.* utterance/speaker, 214–216,
 333–334, 506, 528, 582–585, 646
 externalist theories, 190, 641
 ideational theories, 299–301, 771
 implicit/explicit, 215
 of language, 383
 as mental state *see* Representation
- metaphorical, 445, 448
- Millian theories, 42, 143, 610, 612
 natural *vs.* non-natural, 42, 43, 216,
 505–506, 663
 overview of philosophical theories, 433–438
 procedural *vs.* conceptual *see* Relevance theory
- proper names, 41, 42, 609, 771
- realist theory, 17, 607
- reference, 41, 47, 53
- speaker, 29–30, 97, 214–216, 773–774
- speech-act-centered, 215–216
- stimulus-meaning, 322, 323, 631
- use theories *see* Use theories of meaning
- utterance, 215, 506, 691–692
- verificationist theory *see* Verificationism
- see also* Ambiguity; Context; Context principle;
 Definition; Descriptivism; Literal meaning;
 Meaning; Propositions; Reference; Sense
 and reference; Synonymy; Truth; Truth
 conditions; Use theories of meaning
- Meaningfulness Problem, 186–187, 188
- Measurement
 comparatives, 69
 non-count nouns, 419
- Media discourse
 relevance theory, 654–655
- Mediative *see* Evidentiality
- Medieval Hebrew *see* Hebrew
- Meinong, Alexis
 fields of work
 existence, 213, 228, 636
- Mellor, D H, 741
- Mentalese, 53, 442–444, 656, 750
 arguments for, 442
 theories of mental processing, 715–716
- Mentalism, 561
 internalism, 180, 796–797
 semantics and pragmatics, 688, 692
- Mental mapping, metaphor, 453
- Mental models *see* Representation, mental
- Mental representation *see* Representation, mental
- Mention theory, irony, 341
- Mereology, 574, 576
 non-count reference, 417
- Merge (transformation), 597
- Meronymy, semantic relations, 427–428
- Meronymy, 194–195, 359
see also Hyperonymy; Hyponymy; Prototype
 semantics; Synonym(s)
- Meta-ethics, moral objectivity, 539–544
- Metalanguages, 444–445
 definition, 444
 object language, 558, 804
 semantic, 444–445
 tense, 742
- Metalinguistic negation, 517
- Metalinguistics
 counterfactuals, 110
- Metaphor, 445–449, 449–456
 analogy theories, 447–448
 conceptual *see* Conceptual metaphor
- literal language, 447
 non-standard use, 528
 polysemy, 578
 psychological models, 452
 'career of metaphor' theory, 453
 categorization, 453
 class-inclusion model, 453
 cognitive mechanisms, 452–453
 comparison models, 453
 property sharing, 452
 standard view, 450
 cognitive mechanisms, 451
 'constraint satisfaction', 451–452
 Grice, H Paul, 450
 psycholinguists, 451
 psychological tests, 450
 substitution theory, 447–448
 in thought, 453
 mental mapping, 453
 neural mapping, 453
 ubiquity, 449
 underestimation, 450
 understanding, 450
 workers in
 Grice, H Paul, 450
see also Compositionality; Lexical semantics;
 Semantic change
- Metaphysics, 71–72
 facts, 456–457, 463
 ineffability, 365
 logical positivism, 810, 812
 MCT properties, 457, 460, 463
 substitution salva veritate, 457, 460, 463, 464
 time and tense, 741, 742–743
see also Realism
- Metaphysics, 26
- Metasemantics
 context principle, 88, 91–92
 representation, 180–181, 729
- Metonymy, 528
 polysemy, 578
 workers in
 Johnson, M, 236
 Lakoff, G, 236
see also Cognitive semantics; Metaphor; Polysemy
- Mill, John Stuart, 157, 609, 771
- Miller, George, 51
- Millianism
 direct reference theories, 42, 143, 158, 610,
 612, 771
 No Proposition View, 186
 proper names, 42, 158, 610, 612
- Millikan, R, 508, 509, 642
- Mimesis, 197
- Mind
 computational and representational theories,
 443, 567–568
 grammar, 25
 innateness, 327, 328
 mind-body problem, 222, 748
 philosophy of, 508
see also Cognitivism; Language-thought
 relationship; Mentalese; Representation,
 mental

- Minimal covers of sets, 575
 Minimalism, 207, 261, 596–597
 reference, 643
 semantic, 24, 691, 694
 truth, 142, 776–777, 779
 see also Deflationism; Government-binding theory; Transformational grammar; Transformations
 Minimalist Program, history/development, 791
 Minimal Link Condition, 602
 Minsky, Marvin, 51
 Misidentification
 immunity to error (IEM), 305–308
 openness to error (OEM), 305–308
 Misrepresentation, 506, 657
 Modal auxiliaries, default semantics, 130
 Modality
 analyticity, 8–9, 96–97
 count and non-count nouns, 419
 de dicto, 10, 125, 127
 de re, 126, 127
 modal base, 82–83
 natural kind terms, 502, 503
 possible worlds theories, 581, 582, 794
 Modal logic, 221, 465–472
 definition, 465–466
 K system, 466
 T system, 467
 see also Conditionals; *De dicto/de re* vs.; Propositional attitudes; Tense logic
 Modal operators, 465–466, 739
 Modal verbs *see* Mood/modality
 Model of the fundamental factors of communication, 61*f*
 Modes of presentation
 empty names, 187, 188
 senses as, 216, 336–337, 609–610, 614, 619, 620, 696
 Modularity, 480–482
 central cognitive systems, 481
 Chomskyan ('C-') modules, 480, 562, 563–564
 domains of competence, 480
 information flow, 480
 innate knowledge, 480
 cognitive architecture, 481
 computational/performance ('P-') modules, 480
 databases, 480
 domain specificity, 480
 mandatoriness, 480–481
 domain specificity, 481
 double dissociations, 481
 Fodorian view, 480–481, 562
 input systems, 481
 mental modules, 480
 mind reading, 481
 output systems, 481
 pragmatic capacity, 481
 relevance theory, 653
 representational/computational modules, 331–332, 734
 Monadic operator, 465–466
 Monomorphemic words, in isolating language *see* Chinese
 Monotonicity, 482–485, 515–516, 524, 626, 681–682
 cautious, 526
 definition, 482–483
 see also Compositionality; Inference; Logic; Polarity items
 Montague, Richard, 263, 486, 681, 697
 fields of work
 binding scope, 677
 compositionality, 75
 PTQ, 492
 Montague Grammar, 494
 relational terms, 293–294
 Montague semantics, 485–496
 definition, 486
 history, 485
 see also Compositionality; Dynamic semantics; Extensionality; Intentionality/intention; Quantifiers
 Mood, clause types, and illocutionary force, 496–500
 Mood/modality
 tense interaction, 276
 Moore, G E, 7, 30–30, 142–143
 Moral discourse
 lying, honesty and promising, 413–415
 objectivity *see* Objectivity in moral discourse
 Moral sense theory, 183
 Morphology, 368, 382
 knowledge of, 387
 of Mentalese, 442
 see also Syntax
 Morphosyntax
 adjectives *see* Adjectives
 numerals *see* Numerals
 workers in
 Trager, George, 786–787
 Morris, Charles, 688
 fields of work
 pragmatics, 792–793
 Movement (syntactic) *see* Syntactic movement
 Müller, Friedrich Max
 comparative philology, 473
 Multivalued logic, 513
 see also Vagueness
 Mutual exclusivity constraint
 lexical meaning acquisition, 430
 Mutual parallel adjustments, relevance theory, 651
- N**
- Names
 empty, 48, 49, 56
 fiction, 185
 proper *see* Proper names
 see also Empty names; Noun(s); Proper names
 Narratology
 Saussurian linguistics, 673–674
 Nativism, 561–562
 concept, 80, 327, 329–330, 443
 knowledge, 331
 Naturalism, 381, 507–509, 561–562, 563–564
 epistemology, 11
 meaning, 42, 43, 663
 mind, 656
 normativity, 507, 508, 509, 531
 ontological, 541
 radical internalism, 563
 reference, 42, 506, 642, 656
 see also Empiricism
 Natural kind terms, 41, 501–505
 causal theories, 501
 descriptivism, 501
 externalism, 503
 innateness, 329
 medieval nominalism, 523
 proper names, 502, 641
 as rigid designators, 502, 504, 660
 Natural language semantics
 extensionality, 220
 intensionality, 220
 metalanguage, 445
 Natural Partitions Hypothesis, 200
 N-bar anaphora, 108
 Neale, S, 20, 154, 461, 647–648
 Neander, K, 509
 Necessary/contingent distinction
 ~a priori/a posteriori distinction, 46
 two-dimensional semantics, 796
 Negation, 279–281
 Boolean algebra, 37, 254
 comparatives, 70
 discourse representation theory, 163
 irony, 341
 monotonicity, 483
 negative-raising, 130
 philosophical aspects, 510–513
 classical, 510–511, 511–512
 De Morgan/Ockham, 510–511, 512
 intuitionistic, 510–511, 512
 minimal, 511, 512
 presuppositions, 589, 591–592
 radical, 594, 595
 scope, 128
 semantic aspects, 513–519
 see also Denial
 Negative polarity terms, 36, 37, 515, 623–624
 definition, 390
 licensing conditions, 256
 presuppositions, 592–593
 scope, 484
 Negative-raising, 130
 Nell, Alexander Melville, 229
 Neo-Davidsonian analysis
 event arguments, 194
 nominalization, 194
 see also Davidson, Donald; Event-based semantics
 Neo-Fregianism, 158, 536
 Neogrammarianism
 history of, 472–473
 Neo-Gricean pragmatics
 cooperative principle, 104
 Neo-Millianism, 188
 Neo-performative hypothesis, syntax-pragmatics interface, 719, 719*f*, 721*f*
 Neo-Platonism
 the ineffable, 365
 Neural mapping, metaphor, 453
 Neuroscience, 567–568
 Newmeyer, F J, 207, 269
 Nida, Eugene Albert, 786–787
 Nietzsche, Friedrich, 446
 Nihilism, 809
 Nominalism, 520–524
 moderate realism, 523
 syncategoremata, 521, 715–716, 717
 types, 799
 universals, 370, 520, 521–522, 523, 635
 Nominalization, 194
 transformational grammar, 762
 Nonce uses, 528, 529
 Noncognitivism, 542
 Non-communicated acts, relevance theory, 652
 Noncompositionality, 425
 Noncount nouns *see* Mass nouns
 Nonhuman language, history, 478
 Nonlinguistic knowledge *see* Encyclopedic knowledge
 Nonliterality of sentences, 691
 Nonmonotonicity *see* Inference; Logic
 Nonsegmental phoneme *see* Phoneme
 Non-standard quantifiers, 626
 Non-subject noun phrases, 626
 Non-systematic polysemy, 577
 Non-thematic subjects *see* Pleonastic subjects
 No Proposition View, 186, 187
 Normativity, 530–533
 conversational, 29–30
 lying, honesty and promising, 413–415
 naturalism, 507, 508, 509, 531
 see also Rules
 Norms, 69
 Noun(s), 417–421
 compound
 Boolean compounds, 625
 concrete, 418, 419
 coordinate noun phrases, 576
 count, 36, 37, 417
 natural kind terms, 501
 non-count nouns, 419
 plural, 418, 419
 equi noun phrase deletion, 764
 events, 38
 indexicality, 325–326
 mass nouns *see* Mass nouns
 mass terms, 38
 non-count, 417, 419
 as objects, 626
 plural, 417, 418–419
 proper names *see* Proper names
 Noun phrases *see* Negative polarity terms
 Noun phrases
 bare, 135
 definite/indefinite noun phrases, 131–138
 definition, 388
 deletion, 758–759, 764

equi noun phrase deletion, 758–759
 quantified nou, 623–629
 Nudity, semiotics of, 62
 Numerals
 default semantics, 130

O

Oaklander, L N, 741, 742
 Object(s)
 abstract, 228, 545–546, 798, 799
 individuation, 418–419
 kinds, 545–546
 negative objectual, 213
 object-dependent thoughts *see* Object-dependence
 properties, 212, 228, 545–547
 vague, 807–808
 Object-dependence, 535–539
 senses, 698
 singular thoughts, 537
 Objectivity in moral discourse, 539–544
 cognitivism, 539, 540, 542
 convergence, 539, 543
 independence, 539, 540, 542, 543
 methodological, 539–540
 ontologica, 539–540, 543
 truth, 539, 540–541, 542, 543–544
 Object language, 444–445
 definition, 444
 metalanguage, 558, 804
 Object noun phrases, 626
 Ockham, William, 522, 715–716, 717
 Okanoya, K, 205
 On interpretation, 27, 585
 Ontological categories *see* Semantic categories
 Ontology, 560, 562–563
 logic, 401–402, 405
 moral objectivity, 539–540, 543
 ontological pluralism, 562–563
 realism, 522, 523
 syncategoremata, 716–717
 thought, 747
 see also Existence
 On What Is Not (Gorgias), 569–570
 Opacity problem, 613–614
 see also Frege's problem
 Operators
 logic *see* Logic, operators
 modal, 45, 125, 153, 681
 sentential, 124, 125, 126
 Opposition *see* Antonyms/antonymy
 Optimality Theory, 382, 383
 Lexical Functional Grammar *see* Lexical Functional Grammar (LFG)
 Opting out, cooperative principle, 101
 Optional arguments (semantics), 194
 'Or' (conjunction), 254–255
 Ordering relations, 69
 Ordinary language philosophy, 55, 548–551
 behaviorism, 33, 34–35
 epistemology, 189
 logical form, 412
 see also Austin, John Langshaw; Grice, Herbert Paul; Ryle, Gilbert; Strawson, Peter Frederick
 'Original word game' (Brown 1958), 428–429
 Origin and evolution of language, 551–556
 19th Century, 555
 20th Century, 555
 classical antiquity, 551
 enlightenment, 552
 experiments in, 551–552
 history, 478
 workers in
 Whitney, William Dwight, 478–479
 see also Animal communication; Apes, gesture communication; Bacon, Francis, Lord Verulam; Bopp, Franz; Cognition; Condillac, Etienne Bonnot de; Grammaticalization; Herder; Johann Gottfried; Jespersen, Otto; Locke, John; Müller, Friedrich Max; Plato; Rousseau,

Jean-Jacques; Vico, Giambattista;
 Whitney, William Dwight
 Ostension, 140
 OT-LFG *see* Lexical Functional Grammar (LFG);
 Optimality Theory

P

Palmer, F R, 497
 Papineau, D, 509
 Paradigm(s)
 scientific, 66–67
 Paradoxes
 semantic, 557–559, 804
 Berry's Paradox, 367, 558
 Grelling-Nelson Paradox, 557–558
 Liar Paradox, 511, 557, 776
 Richard's Paradox, 558
 Strengthened Liar Paradox, 557, 559*t*, 774, 776
 sorites, 807, 808
 Paradox of the Horned Man, 591
 Paraphrase
 nominalism, 523
 see also Synonym(s)
 Parmenides, 569
 empty names, 186
 Parole *vs.* langue, 345–346
 Parsimony of levels, 128–129
 Parsing, 22, 25, 181–182, 631
 Parsons, Talcott, 20, 228
 Partee, Barbara H
 Rooth, M, association with, 682
 Partial identity, 165
 Partially ordered sets, 36
 Partial order relations, 36, 38
 Particles *see* Syncategoremata/categoremata
 Partonomy, semantic relations, 427–428
 Parts and wholes, 71–72, 194–195, 359
 Passage of time, 741, 743
 Passives
 transformation, 754
 Pastness, 741
 Paul of Gelfria, 716
 Pavlov, I, 32
 PC *see* Compositionality
 Peacocke, C, 16–17
 Pedagogy
 cooperative principle, 105
 Peirce, Charles Sanders, 62, 324–325, 778, 797, 799
 fields of work
 communication, semiotic approaches to, 62, 62*f*
 Pelletier, F J, 74
 P-entailments, 514, 589–596
 definition, 589
 Perception
 empiricism, 183, 185, 299, 300
 tacit knowledge, 735–736
 Percepts, 200
 Perdurance, diachronic identity, 303–304
 (present) perfect tense *see* Tense
 Performance
 competence, 24–25, 688, 693
 Performatives
 clauses, 498
 sentences, 548, 801–802
 speech acts *see* Speech acts
 Perlmutter, David M
 fields of work
 relational grammar, 790
 Perlocutionary speech acts, 437, 548, 707–708
 Perry, John, 158, 159, 191, 326, 700
 Barwise, J, association with, 461, 463–464
 Personal common ground, 86
 Personality types, 198
 Peter of Ailly, 716
 Peter of Spain, 715, 716–717
 'Pet fish' problem, 77
 Phenomenalism, 560–561, 636–637, 810
 Philosophy, 243
 comparative *see* Comparative philology

Philosophy
 analytic, 6–9
 cognitive science, 47–56
 as first philosophy, 47
 pragmatic acts *see* Pragmatic acts
 see also Ideal language philosophy; Ordinary language philosophy
 Philosophy of linguistics, 53–54, 560–565, 566–568, 704–705
 anti-functionalism, 561
 behaviorism, 561
 Cartesian doctrine, 561–562
 central processing system, 562
 Chomsky, 561
 competence, 561
 constructivism, 562
 E-language, 561
 empiricism, 561–562, 563–564
 encapsulation, 562
 epistemology, 560
 fallibilism, 560
 falsificationism, 560
 Frames of Reference, 564
 generative linguistics, 562–563
 generic conception of language, 563–564, 564–565
 hypothetico-deductive method, 560
 I-language, 561, 563–564
 innateness hypothesis, 561–562
 instantiation, 563–564
 instrumentalism, 560–561
 internalism, 561, 563
 radical internalism, 563
 intuition, 562–563
 knowledge of language, 562–563
 language, 563–564
 language acquisition, 561–562
 language of thought (LoT), 562, 563–564
 linguistic knowledge, 562–563
 internalized knowledge, 563
 mutual knowledge, 563
 phonological knowledge, 563
 logical positivism, 560, 561
 mentalism, 561
 modularity, 562, 563–564
 emergent modularity, 562
 nativism, 561–562
 naturalism, 561–562, 563–564
 radical internalism, 563
 nature and culture, 565
 ontology, 560, 562–563
 ontological pluralism, 562–563
 performance, 561
 phenomenalism, 560–561
 philosophy of science, 560
 physical representation, 563–564
 Platonism, 562–563
 Popper, 560
 psychologism, 561, 562–563
 rationalism, 561–562, 563–564
 realism
 abstract, 562–563
 scientific, 560–561
 Sapir-Whorf hypothesis, 564
 scientific method, 560
 scientific realism, 560–561
 verificationism, 560
 see also Chomsky, Noam
 Philosophy of mind, 508
 Philosophy of science, 560, 566–568, 704–705
 Phoneme(s)
 Saussurian linguistics, 672
 Phonetic(s), 382
 linguistic reality, 368
 see also Phonology
 Phonetic underspecification *see* Underspecification
 Phonological change
 homonymy, 577
 Phonology, 22, 23–24, 24–25, 368, 382
 change *see* Phonological change
 feature, 383
 generative *see* Generative phonology
 knowledge of, 387

- Phonology (*continued*)
 phonemes *see* Phonemes
 Prague School, 783
 Twaddell, William Freeman, 786–787
- Phrase(s)
 denoting, 152, 153
 structure *see* Phrase structure (constituency)
- Phrase structure (constituency)
 representation, 597
 structure *see* c-structure
- Physicalism, 613, 636, 810
- Piagetian conversation task, 430
- Pidgins, 208
- Piedmontese** *see* **Italian**
- Pietroski, P M, 4, 181, 338
 co-workers and associated workers
 Crain, S, 332
 Gualmini, A, 332
- Pike, Kenneth Lee
 fields of work
 tagmemics, 787
- Pinker, S, 204–205, 207, 209, 329, 332
- Plantinga, A, 580, 581
- Plato, 552*f*, 570
Cratylus, 570, 571–574
 distinctions of meaning, 570
 epistemology, 183, 330
 the ineffable, 365–366
 origin of language, 552
 reference, 637
 sentences, 28
Sophist, *The*, 26–27
 universals, 523
- Platonism, 562–563
- Platts, M, 334
- Pleonastic subjects, 587
- Plurality, 164, 574–577
 definite noun phrases, 624
 definition, 574
 indefinites, 678
 semantics, 195
 uniqueness, 132
see also Conjunction; Coordination; Mass terms
- Poincaré, H, 2
- Polarity items, 515
 negation, 484
 reversal, 519
see also Indefinite(s); Monotonicity
- Politeness, 531
 cooperative principle, 104
 relevance theory, 654–655
- Pollard, C, 262
- Polysemy, 202, 577–580
 definition, 577
 lexical semantics, 358
see also Metonymy
- Popper, K R, 373, 373*f*, 375, 560, 562–563, 811
- Positive polarity items, 515
- Possessive(s), 424
 default semantics, 130
- Possessive pronouns, 425
 semantic relations, 427–428
- Possible futures, 739
- Possible worlds, 8–9, 227–228, 447, 580–582
 abstractionist, 581
 concretist, 581
 counterpart theory, 581
 modality, 581, 582, 794
 propositions, 618, 620
 rigid designation, 659, 660
 semantics, 82, 171, 447, 489, 618
 counterfactuals, 110, 111
 modal logic, 466, 470
 update semantics, 175
see also Discourse
- Postal, Paul M, 788
 fields of work
 generative semantics, 790
 relational grammar, 790
- Poverty of stimulus, 33, 331
- PPC/s13/s0, 593
see also Predicate logic
- PPIs *see* Positive polarity items
- Pragmatic determinants, 99, 215, 582–585, 690, 692
 attitude ascriptions, 615–616
 indexicality, 582–583
 Linguistic Direction Principle, 691
 narrowing, 215
 unarticulated constituents, 583
- Pragmatic indexing *see* Indexicality
- Pragmatics, 53–54, 308–321, 396, 795–796
 anaphora, 144, 352
 appropriateness, 32, 33
 Availability Principle, 690
 bestrangement, 319
 centrality of, 720
 constraint *see* Constraint(s)
 contexts, 310
 development *see* Pragmatics development
 Functional Independence Principle, 690
 grammar, 24
 Gricean, 615–616, 688–689, 690, 691
 history *see* Pragmatics, history of
 linguistic reality, 368
 local *vs.* global processes, 528, 690–691, 692
 media *see* Media discourse
 metaphor, 447
 modular concept, 688
 negation *see* Negation
 neo-Gricean *see* Neo-Gricean pragmatics
 origins/development, 792
 Radical, 616
 Scope Test, 690
 semantics *see* Semantic(s), pragmatics and
 situations, 310
 definition, 310
 social conventions, 311
 speech, 314
 syntax *vs.*, 382
 transfer, 693
 truth theory, 142, 778
 understanding/misunderstanding, 310
 Ranciére, Jacques, 310
 workers in
 Austin, John Langshaw, 312–313, 792–793
 Bakhtin, Mikhail Mikhailovich, 317
 Carston, R, 690, 693, 694
 Grice, Herbert Paul, 792–793
 Lakoff, George, 718
 Morris, Charles, 792–793
 Ranciére, Jacques, 310
 Searle, John R, 313
 Voloshinov, Valentin Nikolaevich, 317
see also Conversation; Discourse
- Pragmatics, history of, 478
 Bakhtin, Mikhail, M, 317
 philosophy, 317
 Voloshinov, Valentin, 317
- Pragmatics development
 conversational skills *see* Conversation, child
 development
 lexical meaning, 429
 ‘Pragmatic strengthening,’ future tense, 277
- Prague School, 783
 history, 783
 phonology, 783
 Saussurian linguistics, 474–475
see also Jespersen, Otto; Mathesius, Vilem;
 Trubetskoy, Nikolai
- Predicate logic
 donkey anaphora, 169
 events, 192–196
 modal operators, 469
 stage-level predicates, 193
see also Aristotelian-Abelardian predicate
 calculus; PPC/s13/s0
- Predication, 585–589
 adjectival *see* Predication, nonverbal
 crosslinguistic studies, 587–588
 definition, 585
 reference, 546
- Predication, nonverbal
 adjectives, 586
- Predication rule, 249–251
- Preference organization *see* Conversation analysis
- Prescription *see* Description and prescription
- Presentism, 636–637, 743
- Presentness, 741
- Present perfect tense *see* Tense
- Presumptive meanings, 128
 implicature *see* Implicature
- Presupposition, 29, 108, 128, 589–596
 definition, 589
 dynamic semantics, 176
 failure, 514
 formal semantics, 247, 252–253
 negation, 517
 projection problem, 590
see also Common ground; Discourse;
 Multivalued logic
- Presuppositional propositional calculus, trivalent
see PPC/s13/s0
- Price, H H, 507
- Priestley, Joseph, 148–149
- Primary intention, 128–129
- Primitives
 conceptual, 353, 355
 Primitive sentences, 76
- Primitivism, truth, 780
- Prince, E, 688
- Prince, E F, 135
- Principia mathematica (1910), 466
- Principle of charity, 67, 322
 radical interpretation, 631, 632,
 633–634, 772–773
- Principle of compositionality (PC), 75
- Principle of humanity, 633–634
- Principles and Parameters Model, 24, 596–605
see also Binding theory; Minimalism; X-bar
 theory
- Principles of phonology* (Trubetskoy), 783
- Prior, A N, 580, 741
- Prior, Arthur, 468, 737
- Prior analytics, 26
- Priorean tense logic *see* Tense logic
- Priority scope, 279–281
- Priscianus Caesariensis
 publications
Institutiones grammaticae, 714, 715
- Private language argument, 435, 548, 605–608
 Cartesianism, 606, 607
 communication, 66, 301
 Wittgenstein, 301, 436, 550, 605, 606, 750–751
- Probabilistic theories of concepts, 80
- Probability, conditional, 83
- Procedural analysis, discourse markers, 441
- Procedural encoding, relevance theory, 651
- Procedural meanings
 conceptual *vs.*, 438–442
 conventional implicature, 440
see also Discourse
 development, 439
- Process of recovery, anaphora, 144
- Prodicus, 569
- Productivity
 of language *see* Creativity
 of thought, 79, 444
- Prohibitive negation *see* Negation
- Projection problem, 590
see also Anaphora; Discourse; Presupposition
- Promising, 413–415
 expectation theory, 413–414
 non-informational theory, 414, 415
 practice theory, 413–414
- Pronouns, 134, 247
 anaphora, 19–21, 145, 160, 161, 326, 592, 763
 binding constraints, 603
 demonstrative
 character/content distinction, 44, 216
 reference, 324
 indexical
 character/content distinction, 44, 216
 reference, 157, 324, 637, 640–641, 643
 personal, 324, 325
 possessive *see* Possessive pronouns
- Pronunciation
 knowledge of, 387
- Proper names, 134, 609–613, 637
 causal-historical theory of reference, 41, 42,
 159, 641

- description theories, 638, 660
 direct reference theory, 433, 435–436, 612, 639
 fiction, 157, 228
 hybrid theories, 641
 indexicality, 325–326
 individuals, 435–436
 meaning, 41, 42, 609, 771
 Millian theories, 42, 158, 610, 612, 771
 natural kind terms, 502, 641
 object-dependent thoughts, 535–536
 reference, 41, 139–140, 609
 as rigid designators, 139–140, 659, 660, 742
see also Empty names
- Properties**
 existence as, 212, 214
 functions, 547
 modal-causal-temporal, 457
 objects, 212, 228, 545–547
 relational *vs.* intrinsic, 222
- Proper treatment of quantification *see* PTQ
- Property sharing, metaphor, 452
- Proportional determiners, 625
- Propositional attitudes, 166
see also De dicto/de re *vs.*; Default(s)
- Propositional calculus, 221
 default semantics, 129
 Russellian structured propositions, 172
see also Tense logic
- Propositions**, 617–622
~a priori/a posteriori distinctio, 17, 96–97
 assertions, 28, 29
 Fregean approach, 620
 gappy, 186, 188
 general, 157
 logical consequences, 409
 Millian approach, 186
 minimal, 691
 necessary/contingent distinction, 549
 ontology *vs.* semantics, 620–621
 primacy, 93
 propositional functions, 402
 Russellian approach, 42, 124, 402, 404, 619, 620
 singular, 157, 158, 186, 187, 325
 slingshot argument, 457, 462–463, 464
 as structured, 186, 187
 thought, 613–614
 truth-value, 774, 775, 776, 777–778
 unity, 621
 workers in
 Boole, 399
 Leibniz, 398–399
 Wittgenstein, 366, 403
- Propositive clause, 498
- Prosody**
 grammatical meaning, 294
see also Animal communication; Phonemes; Phonology; Universals
- Protogoras, 497, 569
- Proto-speech-acts, 335
- Prototype semantics
 ‘pet fish’ problem, 77
see also Color; Polysemy;
see also Stereotype semantics
- Psycholinguistics, 566
 irony, 342
 lexical acquisition tests, 428, 430
 productivity *see* Productivity
- Psycholinguists, metaphor, 451
- Psychological tests, metaphor, 450
- Psychologism, 301, 561, 562–563, 618
- Psychology**
 formal-language theory, 22
 innateness, 330, 331
 linguistics as, 179, 566, 567
 meaning, 89, 93
- PTQ, 492
- Public participation, media discourse *see* Media discourse
- Pufendorf, Samuel, 554
- Pustejovsky, J., 362, 694
- Putnam, Hilary, 51
 fields of work
 analytic *a posteriori* claims, 3
- belief revision, 14
 causal theories, 503, 504
 character/content distinction, 773
 communication, 189–190
 definition, 140
 descriptivism, 502–503, 504, 640–641
 externalism, 179, 222–223, 637
 natural kind terms, 223, 503, 612
 sense and reference, 697, 698
 Twin Earth experiment, 222–223
 Pylyshyn, Zenon, 326–327
- Q**
- Qualia structure, 355, 362
- Qualitative productivity *see* Productivity
- Quantification**
 adverbs, 5
 anaphoric pronouns, 19, 20
de dicto/de re, 126
 descriptions, 151–152, 155, 776
 existence, 6–7, 211, 213, 717
 generalized, 681–682
 generic reference, 291
 interrogatives, 686
 logic, 399, 400–401, 402
 logical form, 411–412
 non-count nouns, 417, 419
 restricted/unrestricted, 153
 workers in
 Frege, 716
- Quantifier domain restriction, 692–693
- Quantifiers**, 76, 279
 binding scope, 678
 comparatives, 70
 coordination, 484
 decreasing generalized, 624
 definition, 623
 determiners, 161
 existential
 binding scope, 193
 generalized, 36, 164, 482–485, 494–495
 generic, binding scope, 193–194
 locative existentials, 135
 negation, 483
 scope, 482–485
 scope ambiguity, 193, 627
 semantic properties, 623–629
 universal, 625
- Quantifying-in rule, 160–161
- Quantitative productivity *see* Productivity
- Quantity**
 continuous/discrete, 419
- Qua-problem, 504
- Quasi-indicators, 192
- Question(s)**
 processing, cooperative principle, 105
- Quine, Willard van Ormand, 169, 262–263
~a priori knowledge, 11, 12
 fields of work
 analycity, 2, 8–9, 14, 17, 78–79, 97, 182, 184, 812
 attitude ascriptions, 614–615
 Belief-Revisability Argument, 14–15
 communication, 67
 confirmation holism, 14, 34, 183–184
 context principle, 88, 775
 conventionalism, 97
 definition, 140–141
 deflationism, 142
 descriptive theory of meaning, 189–190
 ideal language philosophy, 47
 identity, 302–303
 indeterminacy, 52–53, 67, 322, 508, 642, 772
 innateness, 328
 mass and count nouns, 418–419
 meaning skepticism, 436, 618, 631, 772
 naturalized epistemology, 507
 nominalism, 523, 635
 proper names, 54–55
 propositions as abstract entities, 618
 radical translation, 322, 631, 632
 reference skepticism, 642
- semantic indeterminacy, 52–53
 truth, 779
 type/token distinction, 799
 verificationism, 810, 812
- Quine–Duhem thesis, 14, 295, 812
- Quintilian, Marcus Fabius
 fields of work
 metaphor, 447–448
- Quirk, Charles Randolph, 150
- Quote names, 804
- R**
- Radical falsity, 595
- Ramsey, F. P., 7, 142, 779
- Rancière, Jacques
 pragmatics, 310
- Rask, Rasmus Kristian, 373–374
 fields of work
 comparative philology, 473
- Rationalism, 379, 380, 561–562, 563–564
 Chomskyan, 733, 736–737
 data, 379–380
 empiricism, 183, 184, 185
 idealization, 379–380
- Rationality**
 language, 748
 mind, 657–658
- R-drop, 380–381
- Realis/irrealis distinction, 497
- Realism, 17, 381, 607, 716–717
 abstract, 562–563
 antirealism, 634–637
 global, 636–637
 intentionality, 703
 moderate, 520–524
 ontological commitment, 522, 523
 scientific, 560–561
 of types, 799
 universals, 521, 635
 of words, 520
- Reality**
 ultimate, 365–366
- Reasoning**
 abductive, 267
 holistic, 267
 instrumental, 98–99
 logical consequence, 407
 nonmonotonic, 524, 526
 temporal, 265
- Recanati, F.
 fields of work
 communicative intentions, 334
 direct reference, 158
 semantic/pragmatic boundary, 690–691, 692, 693
 truth-conditional semantics, 689, 692
- Reciprocals, 766, 768–769
- Reducibility, compositionality, 71–72
- Reductionism, 636
- Reference**, 637–644
 acquaintance-based theory, 536, 639, 641, 643, 647–648
 anaphoric *see* Anaphoric reference
 basic unit of meaning, 47, 49, 53
 cumulative, 418–419
 direct *see* Direct reference theory
 externalist theories, 42, 641
 identity, 107
 indirect *see* Indirect (mediated) reference
 information-theoretic approach *see* Causal theories of reference
 inscrutability, 508, 642
 intentions, 129
 internalist theories, 223, 641, 643
 meaning, 41, 47, 49, 53, 638, 771
 minimalist approach, 643
 naturalistic theory, 42, 506, 642, 656
 proper names, 41, 139–140, 609
 shift, 614, 615
 teleological theory, 642
 workers in
 Aristotle, 636, 637

Reference (*continued*)

- Carnap, Rudolf, 697
 Chomsky, Noam, 53, 642
 Donnellan, K., 41–42, 159, 640
 Dummett, Michael Anthony Eardley, 698
 Evans, Gareth, 43, 611, 641, 698
 Frege, Gottlob, 41, 157, 433, 435, 609–610, 612, 614–615, 638, 640, 696, 697, 798
 Kaplan, D., 41, 159, 641–642
 Kripke, Saul, 41–42, 158, 159, 435–436, 503, 610, 615–616, 641–642, 697, 698
 Lewis, David, 155, 637, 638, 688
 Millikan, R., 508, 509, 642
 Plato, 637
 Putnam, Hilary, 697, 698
 Russell, 615
see also Causal theories; Descriptivism; Sense and reference
- Referential *vs.* attributive descriptions, 644–648
- Reflexives, 627
 transformations, 764
- Reichenbach, Hans, 324–325
- Reid, Thomas, 744, 745
- Reimer, M., 647–648
- Reiter, R., 525
- Relation
 cognitive, 299–300
 expressive, 299–300
 representational, 334–335, 408
 semantic, 299–300, 334–335
 thematic *see* Thematics, relations
- Relational concepts
 concrete concepts *vs.*, 293
- Relational grammar
 history/development, 790
 Perlmutter, David M., 790
 Postal, Paul M., 790
see also Passives
- Relationality, nouns *see* Noun(s)
- Relationships
 part/whole *see* Meronymy
 semantic *see* Semantic relations/relationships
 between sets, 251
 syntactic *see* Syntactic relationships
see also Equivalence
- Relative clauses
 donkey anaphora, 170
see also Nominalization
- Relative tenses *see* Tense
- Relativism
 epistemic, 95–96
 moral, 542, 543–544
- Relevance assessment, relevance theory *see* Relevance theory
- Relevance theory, 616, 648–655
ad hoc concept formation, 652
 applications, 654
 grammar, 654–655
 humor, 654–655
 literature, 654–655
 media discourse, 654–655
 politeness, 654–655
 translation, 654–655
- as asocial, 653
 code *vs.* inference, 649
 inferential models, 649
 cognitive principle of relevance, 650
 communicated acts, 652
 communicative principle of relevance, 650
 conceptual encoding, 651
 definition, 648
 echo, 652
 empirical evidence for, 653
 explicit *vs.* implicit distinctions, 651
 dual-stage processing, 651
 explicatures, 651
 generalized conversational implicatures, 651
 implicatures, 651
 mutual parallel adjustments, 651
 frame problem, 267, 529
 history/development, 648
 implicature *see* Implicature
 irony, 652
 interpretive terms, 652–653
- modularity, 653
 mutual knowledge *vs.* mutual manifestness, 652
 non-communicated acts, 652
 as post-Gricean theory, 649
 communicative intention, 649–650
 informative intention, 649–650
 procedural encoding, 651
 relevance assessment, 650
 contextual implications, 650–651
 semantics, 439
 Carston, R., 439
 Grice, H Paul, 439
 Sperber, D., 439
 Wilson, D., 439, 648
 workers in
 Carston, R., 439
 Grice, H Paul, 439
 Sperber, D., 439, 648
 Wilson, D., 439
- Representation, 23, 24–25
 linguistic, 656–659
 non-verbal, 365
see also Representation, mental; Use theories of meaning
- Representation, mental, 656–659
 cogency condition, 657–658
 concepts, 78, 184
 grammar, 23, 24–25
 as indirect referent, 614–615
 intention, 334–335, 613, 656, 658
 levels, 411, 412
 private language argument, 605–606, 607
 use/mention distinction, 803–805
 workers in
 Frege, 124
- Representatives, speech acts, 709
- Response differences, speech acts, 712
- RG *see* Relational grammar
- Rhetoric
 workers in
 Aristotle, 780
see also Question(s)
- Rhetorical figures
see Figures of speech
- Rhythm
 word stress
see Word stress
- Richards, Ivor Armstrong, 448
 fields of work
 figurative language, semiotics of, 233
- Richard's Paradox, 558
- Ricoeur, Paul
 figurative language, semiotics of, 235
- Rigid designation, 659–660
 definite descriptions, 659, 660
 direct reference, 158–159
 indexicals and demonstratives, 45, 660, 742
 natural kind terms, 502, 504, 660
 necessary a posteriori truth, 12, 304
 proper names, 139–140, 659, 660, 742
- Robins, Robert Henry, 374, 785
- Role and reference grammar (RRG), 271
see also Aktionsart; Syntactic relationships; Thematics, relations
- Root and affix faithfulness, Optimality Theory, in morphology *see* Optimality Theory
- Rooth, M
 Partee, B., association with, 682
- Rorty, Richard, 9
- Rosch, Eleanor, 79
- Rosenberg, A., 509
- Ross, John R., 381
 fields of work
 generative semantics, 790
 island constraints, 767
 syntax-pragmatics interface, 718, 718f
- Rothstein, S., 587
- Rousseau, Jean-Jacques, 446, 553f
- RRG *see* Role and reference grammar
- Rule(s), 661–664
 acquisition of language, 567
 constitutive, 703–704
 filters, 381–382
 interaction, transformational grammar, 758
- rule-following, 661, 802
see also Normativity
- Russell, Bertrand, 47, 124, 220–221, 466, 469, 514–515
- Berry's Paradox, 558
 fields of work
 abstractionism, 581
 coherence theory of truth, 778
 correspondence theory of truth, 777–778
de dicto/de re, 124, 125, 126
 definites, 131–132
 denotin, 153, 615
 descriptivism, 41, 49, 54, 609–610, 638, 640
 direct reference, 157
 epistemology, 507
 existence, 212, 213
 geometry, 1–2
 inference, 183, 410
 language and social construction, 702
 logical atomism, 8, 402–403, 404
 logic and language, 410, 507–508
 object dependence theory, 536
 propositions, 617, 619, 620
 sense data, 810
 sentence meaning, 90, 124
 singular descriptions, 417, 636
 structured propositions, 172
 tense and time, 741
 theory of descriptions, 49, 54, 591
 theory of proper names, 42, 47, 157, 228–229, 433, 612, 639
 vagueness, 807
 Frege, 697
see also Descriptions, definite; Descriptions, indefinite; Reference
- Russian, 411
- Ryle, Gilbert
 fields of work
 category-mistakes, 8
 ordinary language, 8, 9, 49, 50–51
- Ryle, U., association with Kamp, H, 616

S

- S (logical operator), 739
- Sacks, Harvey, 376
- Sadock, J M
 Zwicky, A., association with, 496–497
- Sag, I A, 262
- Salmon, N., 215–216, 609, 615–616, 647–648, 660
- Sameness *see* Identity
- Same-saying, 339–340
- Sampson, G., 562
- Samuels, R., 328, 331–332
- Sandu, G., association with Hintikka, J., 447
- Sapir, Edward, 785–786
 fields of work
 grammatical meaning, 293
 linguistic relativity, 476
- Sapir-Whorf Hypothesis, 185, 564
- Sarcasm
 irony *vs.*, 341
- Satisfaction conditions, 424, 595
see also Lexical conditions
- Satisfaction rule, 632, 775
- Saussurian linguistics, 474
langue-parole dichotomy, 474
 Prague school, 474–475
 signs *see* Saussurian linguistics, signs
- Saussurian linguistics, signs, 665–676
 concepts and sound patterns, 666f
Cours de linguistique générale, 665
 dialogic theory of the word, 668
 differential signs in narrative, 674
 function of the sign, 666
 Lacan's twin doors, 675f
 Marxism and the philosophy of language, 668–669
 narratology, 673–674
 phonemes, 672
 signification and signals, 667f
 signifié and significant, 675f

- subjectivity in language, 669
 workers in
 Barthes, R, 670–671
 Benveniste, E, 669
 Lacan, 674
 Saussure, Ferdinand de, 566, 782
 fields of work
 choice, 379
 structuralist phonology *see* Prague School;
 Saussurian linguistics; Structuralism
 synchronic perspective, 345, 374
 theory of the sign, 665–676
 publications
 Cours de linguiste générale (CLG), 474
 Ecrits de linguistique générale, 474
 Savage-Rumbaugh, S, 198
 Saying
 implicating *vs.*, 99, 689
 stating *vs.*, 689
 Schiffer, S, 698
 Schlesinger, G, 741
 Schütze, C T, 118–119
 Science
 nature of, 371
 see also Philosophy of science
 Scientific realism, 560–561
 Scope, 279–281
 ambiguities, 411–412
 ambiguity, 168, 193, 493
 de dicto/de re, 125, 126
 definition, 677
 dependency, 677
 donkey anaphora, 170
 independence, 677
 negation, 128, 130, 517–518
 quantifiers, 482–485, 627
 semantic aspects, 677–679
 Scope Test, 690
 Script(s), 690–691
 Searle, John R
 fields of work
 assertions, 28, 227
 causal-historical theory, 43, 641
 description theories, 609–610
 metaphor, 446
 non-intentional Background, 216
 pragmatics, 313
 speech-act theory, 478, 708
 see also Speech act theory
 Second-order logic, 279–281
 Segal, G, 54, 336
 Larson, R K, association with, 54, 339–340
 Segerberg, K, 682–683
 Selectional restrictions, 249–251
 Self, first-person thought, 305–306
 Self-knowledge
 externalism, 190, 525, 537
 tacit knowledge, 733–734
 Self-references
 paradoxes, 367
 Sellars, W
 conceptual role semantics, 658
 language as precondition of thought, 751
 Semantic(s), 53–54, 368
 acquisition *see* Semantic development
 adjectives *see* Adjectives
 ambiguity, 645, 646, 647, 689–690
 anaphora, 19
 Aristotelian, 26–27
 atomist, 296, 297, 420
 autonomy, 350
 categories *see* Semantic categories
 cognitive *see* Cognitive semantics
 compositionality, 72, 335
 computational *see* Computational semantics
 conceptual role, 296, 297, 657, 773
 default, 128–131
 denotation, 220
 denotational grammar, 23
 discourse *see* Discourse
 dynamic, 172–177, 616, 683
 see also Dynamic predicate logic
 event-based, 192–196
 expression-centered, 215–216
 extensional, 337, 682
 feature, 383
 formal *see* Formal semantics
 game-theoretical semantics (GTS), 279–282
 generative *see* Generative semantics
 indeterminacy *see* Indeterminacy
 of interrogatives, 685–687
 interval, 739
 lexical *see* Lexical semantics
 lexical conceptual structure (LCS), 349–356
 metaphor, 448
 molecularism, 296
 natural kind terms, 503
 natural language, 445
 optional arguments, 194
 possible worlds *see* Possible worlds, semantics
 prototype
 ‘pet fish’ problem, 77
 realist *see* Realism
 reference, 157, 159, 637–638
 rules, 16
 semantic empiricism, 184, 185
 semantic innocence, 336–337, 685
 semantic value, 401, 615, 637–638, 680–684
 sentences, 88
 situation, 678, 699
 speech-act-centred, 215–216
 syntax interface, 583, 584
 theoretical constraints, 354
 two-dimensional, 44–45, 46, 794–797
 underspecification, 529, 690, 692
 workers in
 Montague, 485–496
 see also Meaning; Paradoxes, semantic;
 Semantic holism; Truth conditions
 Semantic(s), evolution of, 196–204
 cognitive pre-adaptations, 196
 displacement, 197
 intention-reading skills, 197, 198*t*
 mimesis, 197
 motor evolution, 197
 personality types, 198
 semantic knowledge, 199
 abstract concepts, 202
 cognitive dominance, 200, 200*f*
 color, 199
 completion, 201
 concept-combination, 201
 concept formation, 199
 conceptual integration, 201, 201*f*
 conceptual metaphor theory, 202
 cultural evolution, 203
 definition, 196
 embodiment, 199
 Formal Semantics, 199
 lexical concepts, 196–197, 200, 200*f*, 201
 linguistic dominance, 200, 200*f*
 Natural Partitions Hypothesis, 200
 percepts, 200
 polysemy, 202
 selective projection, 201
 semantic units, 196–197
 see also Descartes, René; Origin and evolution
 of language
 Semantic(s), pragmatics and, 128, 247–248, 254,
 516, 688–695
 attribution *vs.* reference, 644–645
 Availability Principle, 690
 Functional Independence Principle, 690
 mentalism, 692
 minimalism, 691, 694
 pragmatic intrusion, 689–690, 691–692, 693
 Scope Test, 690
 Semantical games, 279
 ‘Semantic bleaching’
 future tense, 277
 Semantic categories, 350, 351, 354–355
 Semantic change, 243
 Semantic competence, 246
 definition, 257–258
 Semantic composition, 252
 Semantic development
 lexical meaning, 427–428
 Semantic field, 448
 Semantic holism, 657
 compositionality, 73, 296, 632–633
 conceptual role semantics, 296
 instability problem, 297
 Semantic modeling, 249
 Semantic opposites *see* Antonyms/antonymy
 Semantic properties, 354–355
 comparatives, 68–71
 quantifiers, 623–629
 truth, 76
 see also Syntactic properties
 Semantic relations/relationships, 248–249,
 427–428, 578
 comparatives, 68–71
 entailment, 482–485, 486
 formal semantics, 246–247, 252–253
 modal logic, 468
 see also P-entailments
 function, 427–428
 inclusion, 427–428
 meronymy, 427–428
 ordering, 69
 partonymy, 427–428
 possession, 427–428
 Semantic type shifting *see* Coercion
 Semiotic triangle, 26
 Sensation, private language argument, 606, 607
 Sense, 220, 221–222
 Sense and reference, 612, 639, 696–699, 771
 context principle, 88
 direct discourse, 435
 Kripke, 158, 435–436
 mode of presentation, 216, 336–337, 609–610,
 614, 619, 620, 696
 substitution principle, 336–337, 464
 Sense datum theories, 8, 507, 549, 810
 Sentence(s)
 action, 3–6
 assertions, 28
 complex, 75–78
 eternal, 775
 identity, 610
 langue vs. parole, 346
 logical consequences, 409
 nonliterality, 691
 non-sentences *see* Subsentences
 observation, 322–323, 631, 632
 verificationism, 295, 772
 as performative, 548, 801–802
 as primary truth-bearers, 635, 774, 776
 context principle, 88–94
 primitive, 76
 reference, 41
 as series of names, 28
 stative, 193
 structure
 representation, 390–391
 subsentences, 91, 498, 499–500, 636
 synthetic/analytic, 10, 13, 14, 16, 17, 404, 405
 types, 27
 utterance *vs.*, 154
 see also Clauses; Grammar; Inflectional phrase
 (IP); Speech acts
 Sentential operations, 465–466
 Sequentiality, 175
 Set theory, 251
 Boolean algebra, 36–39
 minimal covers, 575
 plurality, 574, 576
 Shakespeare, William
 language creativity, 114
 Shoemaker, S, 306
 Signification
 meaning *vs.*, 299–300, 300–301
 Signifie, significant vs., 675*f*
 Similarity, 107–109, 111
 comparatives, 83
 definition, 107
 see also Anaphora
 Similarity comparison processes, 79
 Simile
 metaphor, 446, 447–448
 Simmons, K, association with Blackburn, S, 142
 Simons, P, 798–799

834 **Subject Index**

- Since (logical operator), 739
 Sincerity, assertions, 30–30, 79, 187, 188, 617
 Singular indefinite articles, generic reference, 290–291
 Situational knowledge, 425
 Situation semantics, 678, 699–702
 definition, 699
 see also Discourse; Discourse representation theory
 Situation theory, 699
 Situation type *see* Aktionsart
 Skepticism
 constitutive, 662
 knowledge, 94, 189, 190, 635
 meaning, 508, 618, 631, 662–663, 772
 reference, 508
 thought, 747–748
 Skinner, B F, 31, 33, 52, 750–751, 771
 fields of work
 generativism, 477
 Skolem functions, 76, 279
 Slingshot argument
 facts, 456–457
 MCT properties, 457, 460, 463
 proposition as sentence meaning, 457
 Slips-of-the tongue, 214–215
 Smart, J J C, 741, 742
 Smith, Q, 741
 Smuts, J, 74–75
 Soames, S, 508, 609, 615–616, 647–648
 Social conventions, pragmatics, 311
 Socially recognized function, 426
 Sociolect/social class, 179
 Sociolinguistic(s)
 history of, 478
 irony, 342
 origins/development, 792
 workers in
 Labov, William, 792
 Socrates
 origin of language, 552
 Somatic turn *see* Embodiment
The Sophist, 26–27
 Sophists, 569
 Sorities Paradox, 807, 808
 Space grammar *see* Cognitive grammar
 Spandrels, 205, 207
 Specified Subject Condition, 765–766
 Speech
 pragmatics, 314
 Speech acts, 314, 705–714
 Austin, 705
 illocutionary force indicating device (IFID), 707–708
 illocutionary speech acts, 707
 locutionary speech acts, 707
 perlocutionary speech acts, 707–708
 constatives, 705
 cross-culture variation, 711
 directness, 712–713
 response differences, 712
 uniqueness, 712
 direct, 710
 illocutionary *see* Illocutionary speech acts
 indirect *see* Indirect speech acts
 interlanguage variation, 713
 performatives, 705
 felicity conditions, 706
 Searle, J R, 708
 behabitives, 709
 commisives, 709
 declaratives, 709
 directives, 709
 excitatives, 709
 expositives, 709
 expressives, 709
 felicity conditions, 708
 representatives, 709
 types, 709
 verdictives, 709
 see also Discourse; Sentence(s)
 Speech Act theory, 8, 227, 437, 532
 history of, 478
 moral norms, 413–415
 social construction, 703
 speech act types, 335, 801–802
 workers in
 Austin, John Langshaw, 478
 Grice, Herbert Paul, 478
 Searle, J, 478
 Spell-Out, 599
 Spencer, H, 541
 Sperber, Daniel
 fields of work
 relevance theory, 439, 648
 Wilson, D, association with, 447, 616, 690
 Stage-level predicates, 193
 generic reference, 291–292
 Stainton, R J, 647–648, 685, 686
 Stalnaker, Robert C, 85, 159, 176, 581, 641–642, 682–683, 795–796
 Thomason, R, association with, 5
 Stampe, David, 42
 Standards of comparison, 69
 Stanley, Julia Penelope, 692
 Statement
 truth-values, 774, 776
 States of mind *see* Propositional attitudes
 Stative sentences, 193
 Stern, J, 447
 Stokhof, M J B, 176
 Groenendijk, J, association with, 21, 683–684, 686–687
 Stratification grammar
 history, 787
 Lamb, Sydney M, 787
 Strawson, Peter Frederick, 132
 co-workers and associated workers
 Grice, H Paul, 16
 fields of work
 assertions, 28
 deflationism, 142
 induction, 189
 objects and properties, 545
 ordinary language, 8, 49, 189, 549
 presuppositions, 514, 592
 Russell's definite descriptions, 154, 776, 801
 speaker meaning, 215–216, 334–335
 use theories of meaning, 801
 Structuralism
 linguistic, 566
 post-Bloomfeldian, 786
 Structure
 preservation, 764
 see also Rules
 Studies in the Way of Words (Grice), 100–101
 Subjacency Constraint, 601, 766–767
 Subject
 pleonastic (non-thematic), 587
 Subjectless passives *see* Passives
 Subjunctive mood *see* Hypotheticals
 Subsentences, 91, 498, 499–500, 636
 Subsistence, negative subsistentials, 213
 Substance dualism, 613
 Substitution, 245, 493
 salva veritate, 336, 614, 704
 slingshot argument, 464
 systematicity, 728–729
 Substitutivity condition, 76, 220–221
 Sum operations, 194–195
 Superordination *see* Hyperonymy
 Supervaluation, 514
 Supervaluationism, 808
 Supervenience, 140–141, 222, 730
 Surface underspecification *see* Underspecification
 Süßmilch, Johann Peter, origin of language, 554
 Swadesh, Morris, 786
 Sweden
 Swedish Academy, 113
 Swift, Jonathan, 147
 Symbolic meanings, deixis, 145–146
 Syncategoremata/categoremata, 714–717
 history, 521
 Synecdoche
 Lakoff, G & Johnson, M, 236
 Synonym(s), 244, 359, 487
 see also Hyperonymy; Hyponymy
 Synonymy
 analycity, 14, 15
 compositionality, 74
 definition, 138, 139
 propositional content, 620
 Syntactic analysis
 Aristotelian, 26–27
 Syntactic composition, 252
 Syntactic linkage *see* Clauses
 Syntactic movement
 constraints, 764
 island constraints, 601
 Subjacency Constraint, 601
 transformations, 764
 Syntactic properties
 counterfactuals, 109
 definite/indefinite noun phrases, 134
 see also Semantic properties
 Syntactic relationships
 case *see* Case
 predication, 585–589
 see also Clauses
 Syntactic structures (Chomsky), 753–754, 787–788
 Syntactocentrism, 694, 695
 Syntax, 382
 autonomy
 functionalist view, 269, 271, 272, 273
 binding theory *see* Binding theory
 centrality of, 720
 compositionality, 72
 as computational system, 23, 24
 covert structure, 412
 definition, 249–251
 development *see* Syntax, development
 evolution *see* Syntax, evolution of
 feature, 383
 formal language, 22
 intuitions, 181–182
 lexical meaning development, 429–430
 linguistic reality, 368
 logical, 405
 logical form, 411, 412, 583, 694
 of Mentalese, 442–443
 morphology interface *see* Morphology
 pragmatic determinants, 583
 pragmatics interface, 718–728
 classical variants, 718, 718f, 723
 conflict, 718
 consolidation, 724
 functional categories, 725
 grammar inclusive theory, 726
 HPSG, 720, 721f, 722f
 neo-performative hypothesis, 719, 719f, 721f
 semantics interface, 583, 584
 Katz-Postal Hypothesis, 760
 pragmatics, 688, 694
 vs. pragmatics, 382
 workers in
 Lakoff, George, 718
 Saussure, Ferdinand de *see* Saussure, Ferdinand de
 Trager, George, 786–787
 Syntax, development, 208
 crosslinguistic variation *see* Crosslinguistic studies/variation
 Government-Binding theory *see* Government-Binding theory
 Syntax, evolution of, 204–211
 adaption, 207
 'big picture' and details, 207
 biological or nonbiological evolution, 207
 combinatorality, 206, 207
 cultural evolution, 207
 developments within theories, 208
 emergence, 208
 evolution, 206
 evolutionary stages, 208, 209f
 exaptation, 205, 207
 language, 204
 evidence, 206, 210
 evolution of, 205
 speech, 205
 protolanguage, 208

- psychological phenomenon, 206
 recursion, 206
 syntactocentrism, 206
see also Language of thought; Linguistics;
 Origin and evolution of language;
 Semantic(s), evolution of
- Systematicity, 77, 584, 616, 728–731
 compositionality, 728, 730
 grammar, 729, 730
 of language, 613–614, 616, 729
 processing, 728–729, 730
 representation, 729, 730
 of thought, 444
- Systematic polysemy, 578–579
 definition, 577
- Systemic functional grammar (SFG), 272, 379
 choice, 379
 history/development, 791–792
 register, 379
 stylistic analysis, 379
- ## T
- T (system of modal logic), 467
- Tableaux *see* Optimality Theory
- TAG (tree-adjoining grammar) *see* Tree-adjoining grammar
- Tagmemes
 history, 787
 workers in, Pike, Kenneth Lee, 787
- Tarski, Alfred, 8–9
 fields of work
 deflationism, 142
 Liar Paradox, 558
 truth, 181, 434, 632, 633, 681, 772, 775, 778
- Tautology
 logical necessity, 11, 810
 Wittgenstein, 404, 549
- Teleological theories, of reference, 642
- Telicity, 351
- Template morphology
 Optimality Theory, in morphology *see*
 Optimality Theory
- Temporal adverbial phrases, 193
- Temporal distance, future tense, 277
- Temporal logic *see* Tense logic
- Tense
 evidentiality interaction, 276
 mood/modality interaction, 276
 Tense and time, 682–683, 741–744
 A-theory, 741
 B-theory, 741
 presentism, 636–637, 743
 Tensed-S Condition, 765–766
 Tense logic, 164, 468, 470, 737–741
 definition, 737
 operators, 737–738
 semantics, 738
 see also Event-based semantics; Modal logic
- Term(s)
 singular, 153–154, 157, 325, 535, 635, 660, 696
 substitutivity, 458, 460, 461, 463, 464
 type/token distinction, 798, 799
 universals, 636
- Tesnière, Lucien Valerius, 785
- Testimony, 744–747
 as belief source, 745
 reductionism and anti-reductionism, 746f
- T-glottaling, 380–381
- Thematics
 relations
 predication, 586–587
 semantics, 194–195, 360
- Theology
 negative, 365, 366–367
- Theoretical linguistics, Aristotle, 25–27
- Theory
 default, 525
 deflationary *see* Deflationism
 positivistic, 2
- Theory of descriptions, 591
- Theory of meaning
 cognitive dependency, 424–426
 discourse representation, 128, 160–168,
 171, 173
 see also Conceptual metaphor theory
- Theory of truth, Aristotle, 26
- ‘There’ (existential), 624
- Theta Criterion, 350–351, 354
- Theta-role generalization, 361
- Theta-roles *see* Thematics
- Third Man argument, 520
- Thomas of Cleves, 716
- Thomas of Erfurt, 716–717
- Thomason, R., 74–75
- Stalnaker, R., association with, 5
- Thomasson, A., 229
- Thompson, S. A., 274
- Thought
 abstract *see* Abstraction/abstract thought
 content *see* Externalism; Frame problem;
 Internalism
 first-person, 305–306
 ideational theories, 299, 300
 as inner speech *see* Mentalese
 language *see* Language-thought relationship;
 Mentalese
 language of, 396
 language relationship *see* Language-thought
 relationship
 object-dependent *see* Object-dependence
 productivity, 79, 444
 public language, 185, 656, 750–751, 771
 singular, 535, 537
 as systematic, 444
 see also Language of Thought
- Thought experiments
 examples, 773
- Tirrell, L., 447–448
- T-model, 411
- Tokens *vs.* types *see* Types *vs.* tokens
- Tolerance principle, 404–405, 406
- Tolman, E., 31
- Tomasello, M., 197, 207
- Tooley, M., 741
- Topic-comment analysis, Aristotelian, 26–27
- Trace theory, 767
 definition, 768
 transformations, 597
- Traditions, in animals *see* Animal(s)
- Trager, George
 fields of work
 morphology/syntax, 786–787
- Transfer
 pragmatic, 693
- Transformational Grammar (TG),
 752–770
 c-structure, 757–758
 Derivational Complexity Hypothesis, 757
 extended standard theory, 762
 history, 752–753
 Lexicalist Hypothesis, 762
 rule interaction, 758
 wh-movement, 768–769
 see also Government-binding theory
- Transformations
 constraints, 598
 Merge, 597
 ordering rules, 756
 passives, 754
 subtypes, 755
 trace theory, 597
- Translation
 crosslinguistic synonymy, 217–218
 indeterminacy, 67, 322, 437, 508,
 631, 772
 metalanguage, 445
 radical, 322, 631
 relevance theory, 654–655
 see also Interpretation
- Transportation, definite/indefinite noun
 phrases, 133
- Tree-adjoining grammar (TAG), 23, 24
- Triangulation, interpretation, 632
- Tripartite parallel architecture (of language),
 351, 351f
- Trivalent logic, 513
- Trivalent presuppositional propositional calculus
 see PPC/s13/s0
- Trubetskoy, Nikolai, 783
 publications
 Principles of phonology, 783
- Truss, L., 149
- Truth
 a priori, 1, 11–12, 46, 404
 Kant, 10
 logical positivism, 2, 810
 analytic/synthetic, 16, 184, 810
 a posteriori, 10
 conditions *see* Truth conditions
 necessary/contingent, 10, 46, 796
 as normative, 779
 objectivity in moral discourse, 539, 540–541,
 542, 543–544
 paradox, 558
 pluralism, 780
 realism, 636–637
 theories, 777–780
 coherence theory, 142, 778
 deflationary theory, 142, 776–777, 777–780
 disquotationalism, 779
 identity theory, 780
 minimalist, 142, 776–777, 779
 pragmatic theory, 142, 778
 primitivism, 780
 redundancy theory, 779
 as warranted assertibility, 635
 workers in
 Kant, 10
 Wittgenstein, 403–404, 777–778
 Truth conditions, 26, 45–46, 70, 253–254,
 279–281, 437, 770–774
 Aristotle’s theory, 26
 assumed, 216
 attitude ascriptions, 614–615
 conditionals, 82
 context, 94, 95, 583
 counterfactuals, 109
 Davidson’s theory, 76
 definite descriptions, 645
 donkey anaphora, 169
 intention-based semantics, 334–335
 interpreted logical forms, 337
 interrogatives, 685
 intuitions, 181–182, 635
 language conventions, 97
 lexical semantics, 692
 possible worlds theories, 581–582
 semantic holism, 295, 296
 semantics-pragmatics boundary, 689, 691
 subsentences, 302–303, 636
 substitution, 245
 systematicity, 730–731
 tense, 742
 Truthfulness and trust
 convention, 98
 testimony, 745
 Truth functions, theories of meaning, 434
 Truth theories, correspondence theory, 142, 540,
 775, 777–778
 Truth-value
 Frege, 401, 680, 697
 primary bearers, 774–777
 vagueness, 808
 W., 403–404
 Truth Value Problem, 186–187, 188
 Turbulence intensity, speech aerodynamics
 see Speech aerodynamics
- Turner, M., 201
- Twaddell, William Freeman
 fields of work
 phonology, 786–787
- Twin earth argument, 222–223, 503
- Tyler, A., 202
- Typed logic, 252
 non-subject noun phrases, 626
 ‘type mismatch’ problem, 626–627
- Type-level noncompositionality, 425
- Types
 speech acts, 709
 verbs, 249–251

Types *vs.* tokens, 369, 797–800
 classes, 370
 occurrences, 798–799
 tense, 742
 truth values, 775–776
 type identity theories, 798
 word types, 797
 Typicality, 79
 Typological linguistics
 history, 787
 origins/development, 792
 Typology
 future tense, 277
 Tzetal, 564–565

U

U (logical operator), 739
 UG *see* Universal grammar
 Underdetermination
 meaning, 128, 579
 semantic, 692
 of theory, 67, 508, 642
 Underspecification
 reference *see* Reference
 semantic, 529, 690, 691, 692
 Understanding, 73, 89, 607
 systematicity, 728–729, 730
 Uniqueness, 131, 417, 591
 speech acts, 712
 Universal determiners, 626
 Universal Grammar (UG), 51–52, 99, 179,
 283–284, 347, 736–737
 transformations, 765
 Universal quantifier, 625
 Universals, 244, 354
 conceptualism, 520, 799
 linguistics as science, 393–394
 nominalism, 520, 521–522, 523, 635
 Optimality Theory *see* Optimality Theory
 realism, 520, 635
 types *see* Types
 Until (logical operator), 739
 Update semantics
 Update semantics, 175
 Use of language
 non-standard, 527–530
 standardized, 527
 Use theories of meaning, 49, 50, 189, 438,
 658, 801–803
 Davidson, 296, 437
 deflationism, 143, 803
 ordinary language theories, 49, 50
 Russell, 152, 801
 sentence primacy, 91
 Strawson, 801
 Wittgenstein, 91, 436–437, 550, 801, 802
 Use *vs.* mention distinction, 803–805
 Utterance
 context, 582–583, 583–584
 illocutionary point, 497–498, 499–500
 meaning *see* Meaning
 sentence, 154
 sincere assertive, 186–187

V

Vacuous quantification, 193–194
 Vagueness, 513–514, 578
 philosophical aspects, 807–809
 higher-order, 807
 supervaluationist theory, 808
 Valian, V, 118
 Validationality *see* Evidentiality
 Validity, 407
 Valin, R D van Jr, 271–272
 Value *see* Truth
 Van der Sandt, R A, 518

Vanderveken, D, 28
 Van Eijck, J, 166
 Van Inwagen, P, 229, 581
 Van Valin, R D Jr, 271–272
 Variable(s)
 in comparative clauses, 69
 definite/indefinite noun phrases, 133
 events, 192–193, 193–194
 implicit, 192–193
 syntactic *see* Transformational grammar
 see also Vagueness
 Variable binding operators, 627
 Variable embodiment, 199
 Variable weight, word stress
 see Word stress
 Variation of word order *see* Word order
 VdashVdash, 249–251
 Veltman, F J M M, 176
 Venetian *see* Italian
 Verb(s)
 agreement, 600
 auxiliaries, 753–754
 factive, 279–282
 logical properties, 249–251
 modal auxiliaries, 130
 negation, 256
 ‘possession’ predicates, 424
 VP-ellipsis, 108
 see also Verb phrase
 Verb phrase
 definition, 388
 Verdictives, speech acts, 709
 Verificationality *see* Evidentiality
 Verificationism, 560, 810–813
 atomistic, 33, 34
 epistemology and language, 189, 295
 holistic, 34
 meaning, 12, 184, 772
 observation statements, 295, 772
 see also Logical positivism
 Vico, Giambattista
 figurative language, semiotics of, 232
 Vienna Circle, 33, 560, 772, 810
 see also Carnap, Rudolf; Logical positivism;
 Verificationism
 Viewpoint
 see Aspect
 Violation, cooperative principle, 101
 Voloshinov, Valentin Nikolaevich
 fields of work
 pragmatics, 317
 Vygotskii, Lev Semenovich, 751

W

Walton, K, 229
 Wasow, T, 262
 Watson, J B, 31
 Weatherford, J, 203
 Well-formedness
 syntactic definition, 624
 Wenker, Georg
 fields of work
 linguistic geography, 475
 Westerstahl, D, 74
 Wettstein, H, 155
 Wexler, K, 264
 When-clauses, 193–194
 Whitehead, A N, 466
 Whiten, A, 198
 Whitney, William Dwight
 fields of work
 origins of language, 478–479
 wh-movement
 Full Interpretation (FI), 600
 long-distance, 600–601, 766
 transformational grammar, 768–769
 Whole object constraint, lexical meaning
 acquisition, 430

Wholes and parts, 71–72, 88
 Wholism, compositionality, 71–72
 Whorf, Benjamin Lee, 786
 Wilson, G, 4
 William of Sherwood, 715, 716–717
 Williams, Edwin, 586–587
 Williamson, T, 29, 808
 Wilson, D, association with Sperber, D, 447,
 616, 690
 Wilson, Deidre
 relevance theory, 439, 648
 Wilson, G, 20
 Winch, P, 702–703
 Wittgenstein, Ludwig Josef Johann
 fields of work
 analytic philosophy, 7
 context principle, 88, 90–91
 determinacy of meaning, 322
 ideational theories of meaning, 299, 301
 inexpressibility, 365, 366
 language and social construction, 702
 language games, 606, 635, 802
 language rules, 661, 802
 logical atomism, 8, 405
 misidentification error, 307
 ordinary language, 7–8, 49, 189, 548
 picture theory, 366
 private language argument, 301, 436, 550,
 605, 606
 public language, 606, 750–751
 use theory of meaning, 91, 436–437, 550,
 801, 802
 Word(s), 368, 426
 classes *see* Word classes
 formation *see* Word formation
 idea theory, 299, 436
 learning *see* Lexical acquisition
 learning experiments, 428, 430
 maps/representations, 426–427
 meaning
 development *see* Lexical acquisition
 fixed/conventional, 429
 overlapping/ambiguous *see* Lexical
 ambiguity
 nominalist account, 520
 realist account, 520
 types *vs.* tokens, 797
 Word classes
 Aristotelian terms for, 27
 Worden, R P, 207
 Word formation
 coining words, 112–113
 Word order
 grammatical meaning, 294
 Worlds, possible *see* Possible worlds
 Wright, C, 605, 607–608, 663

X

X-bar semantics, 351–352, 353
 X-bar syntax, 351–352, 353
 X-bar theory, 762–763
 see also Agreement; Binding theory;
 Government-binding theory; Principles
 and Parameters Model; Thematics

Y

Yale shooting problem, 526
 Yngve, V H, 373–374, 377

Z

Zalta, E, 228, 229
 Zen (Chan), 365–366, 367
 Zwicky, A
 Sadock, J M, association with, 496–497